ÓU

Óbuda University

# Characterisation of noisy speech channels in 2G and 3G mobile networks

Master Thesis

To obtain the degree of master at the
Instituto Superior de Engenharia do Porto,
public defend on June 27th 2013 by

Bruno Daniel Moreira Leite

Master in Electrical and Computer Engineering
Telecommunications Specialization
Porto, Portugal.

Supervisor:

Dóra Maros (PhD), Óbuda University

Co-supervisor:

Jorge Botelho da Costa Mamede (PhD), School of Engineering - Polytechnic of Porto

Composition of Supervisory Committee:

Peter Mihajlik (PhD), Technical University of Budapest
Tibor Wührl (PhD), Óbuda University
Sándor Gyányi (PhD), Óbuda University

Author email: 1070199@isep.ipp.pt

*To my Grandparents.*

# Acknowledgments

To the opportunity given to me by my supervisor Professor Dóra Maros to work on such an interesting and challenging topic.
Also to my co-supervisor Professor Jorge Botelho da Costa Mamede for being there when I needed.
Not forgetting the academic staff that was involved and, a student (Csaba Sipos) who was more than helpful along the way.
Last but not least, to my family, closest friends and my girlfriend for all the encouragement and support.

# Abstract

As the wireless cellular market reaches competitive levels never seen before, network operators need to focus on maintaining Quality of Service (QoS) a main priority if they wish to attract new subscribers while keeping existing customers satisfied. Speech Quality as perceived by the end user is one major example of a characteristic in constant need of maintenance and improvement.

It is in this topic that this Master Thesis project fits in. Making use of an intrusive method of speech quality evaluation, as a means to further study and characterize the performance of speech codecs in second-generation (2G) and third-generation (3G) technologies. Trying to find further correlation between codecs with similar bit rates, along with the exploration of certain transmission parameters which may aid in the assessment of speech quality.

Due to some limitations concerning the audio analyser equipment that was to be employed, a different system for recording the test samples was sought out. Although the new designed system is not standard, after extensive testing and optimization of the system's parameters, final results were found reliable and satisfactory. Tests include a set of high and low bit rate codecs for both 2G and 3G, where values were compared and analysed, leading to the outcome that 3G speech codecs perform better, under the approximately same conditions, when compared with 2G. Reinforcing the idea that 3G is, with no doubt, the best choice if the costumer looks for the best possible listening speech quality.

Regarding the transmission parameters chosen for the experiment, the Receiver Quality (RxQual) and Received Energy per Chip to the Power Density Ratio ($Ec/N_0$), these were subject to speech quality correlation tests. Final results of RxQual were compared to those of prior studies from different researchers and, are considered to be of important relevance. Leading to the confirmation of RxQual as a reliable indicator of speech quality.
As for $Ec/N_0$, it is not possible to state it as a speech quality indicator however, it

shows clear thresholds for which the MOS values decrease significantly.
The studied transmission parameters show that they can be used not only for network management purposes but, at the same time, give an expected idea to the communications engineer (or technician) of the end-to-end speech quality consequences.

With the conclusion of the work new ideas for future studies come to mind. Considering that the fourth-generation (4G) cellular technologies are now beginning to take an important place in the global market, as the first all-IP network structure, it seems of great relevance that 4G speech quality should be subject of evaluation. Comparing it to 3G, not only in narrowband but also adding wideband scenarios with the most recent standard objective method of speech quality assessment, POLQA. Also, new data found on $Ec/N_0$ tests, justifies further research studies with the intention of validating the assumptions made in this work.

**Keywords:** 2G, 3G, Speech Quality, MOS, PESQ, AQuA, Codecs, Noisy radio channels

# Bevezetés

Mivel a vezeték nélküli mobil piacon eddig soha nem látott szintre emelkedett a versenyhelyzet, a hálózatüzemeltetőknek egyre inkább arra kell összpontosítaniuk, hogy a szolgáltatásminőséget, azaz a Quality of Service-t (QoS) tekintsék az egyik fő prioritásnak, annak érdekében, hogy a meglévő ügyfeleik elégedettségét fenntartsák, és új előfizetőket szerezzenek. A beszéd minősége a végfelhasználó által érzékelt egyik legfontosabb és legjellegzetesebb példája a QoS-nek, amelynek folyamatos karbantartása és javítása szükséges.

A diplomamunka témája a fenti problémakör tárgyalását tűzte ki céljául. A beszéd minőségének objektív és szubjektív értékelési módszerei mellett, a dolgozat tanulmányozza a második (2G) és harmadik generációs (3G) mobil technológiákban alkalmazott beszédkodekeket a QoS szempontjait is figyelembe véve. A dolgozat megpróbálja feltárni a hasonló bitsebességgel, de más elven működő beszéd kodekek QoS jellemzőit az átviteli paraméterek értékelésével.

A tesztek számára rendelkezésre álló audio készülék limitált műszaki lehetőségei miatt, olyan vizsgálati megoldás és tesztrendszer kidolgozására volt szükség, amely a tesztmintákat, a szabványokban megadottól eltérő módon rögzíti, és bár az új kidolgozott tesztrendszer nem szabványos megoldásokra épül, a kapott eredmények a tesztelés és optimalizálás után megbízhatónak és megfelelőnek bizonyultak.
A tesztek során 2G és 3G hálózatban alkalmazott magas és alacsony közel azonos bitsebességű kodekek kerültek összehasonlítása és elemzésre. Az eredmények alapján megállapítható, hogy a 3G-ben alkalmazott beszédátvitel QoS értékei - összehasonlítva az azonos bitsebességű 2G-s megoldásokkal- azonos tesztkörnyezetben jobbak.

A hangminőség vizsgálatok tekintetében két átviteli paraméter került kiválasztásra, a vevő által meghatározott minőségi paraméter (RxQual), valamint a vett jel egy chipre vonatkoztatott energiasűrüség és zaj ($Ec/N_0$) hányadosának értéke. A mért RxQual értékek összehasonlításra kerültek a különböző előzetes kutatások eredményeivel, amelynek eredménye igazolja, hogy az RxQual közvetlen és megbízható mutatója a

beszéd minőségének. A kapott Ec/$N_0$ értékek nem közvetlenül jellemzik a beszéd minőségét, azonban a vizsgálati eredmények egyértelműen bizonyítják, hogy bizonyos küszöbértékek alatt, a szubjektív értékelés MOS értékei lényegesen csökkennek.
A vizsgált átviteli paraméterek azt mutatják, hogy azok nem csak hálózati menedzsment célokra alkalmasak, hanem segítséget adnak a mérnökök számára az end-to-end beszédminőség meghatározásában is.

A dolgozatban felvázolt új ötletek a későbbi tanulmányok esetében is hasznos segítséget nyújthatnak a beszéd Qos jellemzőinek meghatározására. Tekintettel arra, hogy a negyedik generációs mobil technológia most kezd a világban elterjedni, és egyre fontosabb helyet foglal el a globális piacon, a tisztán IP-alapú hálózati struktúra új problémákat vet fel a 4G beszéd minőséget illetően, ami újabb vizsgálatok tárgyát képezheti. A POLQA egy olyan új szabvány, amely alkalmas a 3G beszédminőségi vizsgálatokon túl a 4G-ben alkalmazott beszédátvitel minőségének értékelésére is, Többek között az új POLQA későbbi hatékony alkalmazásához nyújthatnak segítséget a dolgozatban ismertetett vizsgálati módszer Ec/$N_0$ teszt eredményei.

**Kulcsszavak:** 2G, 3G, beszéd minőség, MOS, PESQ, AQuA, beszédkodekek, zajos radiós csatorna

# Resumo

Com o mercado das redes móveis a atingir níveis de competitividade nunca antes vistos, existe a crescente necessidade por parte dos operadores de rede em focar-se na Qualidade de Serviço (QoS) como principal prioridade, no sentido de atrair novos clientes ao mesmo tempo que asseguram a satisfação dos seus actuais assinantes. A percepção da Qualidade de Voz, por parte do utilizador, é apenas um exemplo de uma característica de QoS em constante necessidade de manutenção e melhoramento.

Sendo nesta temática em que se insere a Tese de Mestrado. Aplicando um método intrusivo de avaliação de qualidade de voz, como meio para um estudo mais aprofundado e, ao mesmo tempo, caracterizando o desempenho dos *codecs* de voz para as tecnologias de segunda-geração (2G) e terceira-geração (3G). Investigando nova informação que possa ser retirada da correlação entre *codecs* com *bit rates* semelhantes, juntamente com a exploração de determinados 'parâmetros de transmissão' os quais podem auxiliar na avaliação da qualidade de voz.

Devido a algumas limitações ligadas ao analisador de áudio (requisito neste tipo de aplicações), existiu a necessidade de procurar um sistema distinto para gravação das amostras de teste. Embora o sistema escolhido não seja padronizado para este tipo de ensaios, após vários testes e consequente optimização dos parâmetros do sistema, os resultados finais consideram-se credíveis e satisfatórios.
Os testes efectuados incluem um conjunto de *codecs* de elevado e baixo *bit rate*, onde a comparação e análise dos resultados levam a concluir que *codecs* de voz 3G têm melhor desempenho, sob aproximadamente as mesmas condições, comparativamente com os 2G. Reforçando a ideia generalizada que 3G é, sem dúvida, a melhor escolha se o utilizador procura uma solução superior a nível de qualidade de voz.

No que diz respeito aos parâmetros de transmissão escolhidos para a experiência, RxQual (Qualidade do sinal Recebido pela estação móvel) e $Ec/N_0$ (razão entre Energia por chip e a Densidade Espectral de Potência), estes foram sujeitos a testes de correlação com a qualidade de voz.

Os resultados de RxQual foram sujeitos a comparação com estudos prévios de outros investigadores, confirmando este parâmetro como um indicador de qualidade de voz bastante fiável.
Quanto a $Ec/N_0$, não é possível declará-lo como um indicador de qualidade de voz, no entanto, este demonstra limites claros para os quais os valores de *Mean Opinion Score* (MOS) decrescem significativamente.
Os parâmetros de transmissão estudados demonstram não só que podem ser utilizados com objectivos de gestão de rede mas como também podem fornecer, ao engenheiro (ou técnico), informação relativa ao impacto que poderá existir na qualidade de voz.

Com a finalização deste trabalho é possível constatar que novos estudos devem ser efectuados. Considerando que a tecnologia de quarta-geração (4G) começa agora a dar os seus primeiros passos no mercado das redes móveis, como a primeira com arquitectura de rede totalmente orientada para IP, parece de grande importância que esta tecnologia seja sujeita a avaliação. Comparando-a com 3G, não só para banda-estreita (300 a 3400 Hz) como também para cenários de banda-larga (50 a 7000Hz), aplicando o mais recente método normalizado de avaliação de qualidade de voz, o POLQA. Por fim, também se verifica como pertinente uma continuação do estudo relativo a $Ec/N_0$ a fim de validar as ilações retiradas neste trabalho.

**Keywords:** 2G, 3G, Qualidade de Voz, MOS, PESQ, AQuA, Codecs, Ruído em Canais rádio

# Contents


**1 Introduction** **1**
- 1.1 Context . . . 2
- 1.2 Objectives . . . 2
- 1.3 Development Plan . . . 3
- 1.4 Outline of the Thesis . . . 4

**2 2G and 3G General Overview** **5**

**3 Speech Quality** **9**
- 3.1 Speech Transmission . . . 10
- 3.2 Codecs and Speech Quality Assessment . . . 13
  - 3.2.1 Subjective Evaluation - MOS . . . 14
  - 3.2.2 Perceptual Evaluation of Speech Quality (PESQ) . . . 16
  - 3.2.3 Perceptual Objective Listening Quality Analysis (POLQA) . 20
  - 3.2.4 Audio Quality Analyser (AQuA) . . . 26
- 3.3 Transmission Parameters . . . 29
  - 3.3.1 Bit Error Rate . . . 30
  - 3.3.2 Block Error Rate . . . 31
  - 3.3.3 Receiver Quality . . . 32
  - 3.3.4 Receiver Level . . . 32
  - 3.3.5 Frame Erasure Rate . . . 33
  - 3.3.6 Carrier-to-Interference Ratio . . . 34
  - 3.3.7 Received Energy per Chip to the Power Density Ratio . . . 35

**4 Source Coding in GSM and UMTS** **37**
- 4.1 GSM Codecs . . . 37
  - 4.1.1 Full-Rate codec . . . 38
  - 4.1.2 Half-Rate codec . . . 38
  - 4.1.3 Enhanced Full-Rate codec . . . 40
  - 4.1.4 Processing functions for FR, HR and EFR . . . 41
  - 4.1.5 Adaptive Multi-Rate codec . . . 45

4.2 UMTS Codecs . . . . . 47
4.2.1 AMR-NB and AMR-WB . . . . . 47
4.2.2 Adaptive Multi-Rate-Wideband codec . . . . . 49
4.2.3 Extended Adaptive Multi-Rate Wideband codec . . . . . 50
4.3 Performance Characterization of Speech Quality . . . . . 51

**5 Implementation of the Thesis Project 55**
5.1 Project . . . . . 55
5.1.1 Requirements . . . . . 56
5.1.2 Selection Phase and Laboratory Set-Up . . . . . 57
5.1.3 System Architecture . . . . . 65
5.2 Testing and Evaluation . . . . . 66
5.2.1 R&S CMW500 Configurations . . . . . 67
5.2.2 Calibration of Volume Parameters . . . . . 68
5.2.3 Tests - Clean Speech . . . . . 71
5.2.4 Tests - Error Conditions . . . . . 74
5.2.5 Results and Performance . . . . . 75
5.2.6 Results Discussion . . . . . 83

**6 Conclusions and Future Work 85**

**Annex A. Test results - AQuA PESQ MOS 93**

**Annex B. Test results - AQuA Voice Quality in Percentage 97**

# List of Figures


1.1 Gantt chart regarding the development plan for the Thesis/Dissertation 3

3.1 General Speech Transmission system process . . . . . . . . . . . . . 11
3.2 Speech quality versus bitrate classification of speech codecs . . . . . 12
3.3 PESQ Architecture . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
3.4 Evolution of ITU-T Recommendations for Speech quality Testing (P.86x) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
3.5 Basic POLQA philosophy . . . . . . . . . . . . . . . . . . . . . . . . 22
3.6 AQuA Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
3.7 AQuA general scheme of the quality estimation system for sound signals 28
3.8 Main speech Key Performance Indicators in GSM . . . . . . . . . . . 30
3.9 Illustration of interference received from co-channels - uplink situation 34

4.1 Through-transport of GSM-coded speech in Phase 2+ for mobile-to-mobile connections (TFO) . . . . . . . . . . . . . . . . . . . . . . . 39
4.2 Block diagram of the GSM half rate speech codec . . . . . . . . . . . 40
4.3 Half rate codec - processing functions reference configuration . . . . 42
4.4 AMR channel encoding principle (bit numbers for TCH/F) . . . . . 46
4.5 Mean opinion score (MOS) example with wideband and narrowband AMR . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
4.6 Family of curves for Experiment 1a (Clean Speech in Full Rate) . . . 51
4.7 Family of curves for Experiment 1b (Clean Speech in Half Rate) . . 53

5.1 Project Proposed tests . . . . . . . . . . . . . . . . . . . . . . . . . . 56
5.2 Rohde & Schwarz CMW500 Wideband Radio Communication Tester 58
5.3 Rohde & Schwarz Signal Generator SME 03 (5 kHz to 3.0 GHz) . . 59
5.4 Illustration of the basic system configuration . . . . . . . . . . . . . 61
5.5 Behringer ULTRA-DI Model DI100 (Active Direct Inject Box) . . . 62
5.6 Schematic of connections for Phone, PC and DI-Box . . . . . . . . . 64
5.7 System Architecture . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

5.8 Recording example (for GSM) with volume parameters optimized - Comparison between reference original signal (above) and respective degraded signal (below) . . . . . 69
5.9 Variable delay in UMTS . . . . . 70
5.10 AQuA PESQ MOS values (Clean Speech) . . . . . 72
5.11 AQuA Voice Quality in Percentage (Clean Speech) . . . . . 73
5.12 AQuA PESQ MOS Average values (Clean Speech + Noise Levels) . 76
5.13 Graphic representation of AQuA PESQ MOS Average values (Clean Speech + Noise Levels) . . . . . 76
5.14 AQuA Voice Quality in Percentage, Average values, (Clean Speech + Noise Levels) . . . . . 77
5.15 Graphic representation of AQuA Voice Quality in Percentage, Average values, (Clean Speech + Noise Levels) . . . . . 78
5.16 AQuA PESQ MOS Average values (Clean Speech + Noise Levels) - High Rate codecs . . . . . 79
5.17 AQuA PESQ MOS Average values (Clean Speech + Noise Levels) - Low Rate codecs . . . . . 79
5.18 RxQual in GSM (Clean Speech + Noise Levels) . . . . . 80
5.19 RxQual versus AQuA PESQ MOS - GSM FR codec . . . . . 80
5.20 RxQual versus AQuA PESQ MOS - GSM EFR codec . . . . . 81
5.21 RxQual versus AQuA PESQ MOS - GSM HR codec . . . . . 81
5.22 $Ec/N_0$ (lower/upper values) in UMTS (Clean Speech + Noise Levels) 82
5.23 $Ec/N_0$ versus AQUA PESQ MOS (UMTS codecs) . . . . . 83

6.1 AQuA PESQ MOS values (Pn = -4.5 dBm) . . . . . 93
6.2 AQuA PESQ MOS values (Pn = -3.0 dBm) . . . . . 94
6.3 AQuA PESQ MOS values (Pn = -1.5 dBm) . . . . . 94
6.4 AQuA PESQ MOS values (Pn = -0.5 dBm) . . . . . 95
6.5 AQuA PESQ MOS values (Pn = 2.0 dBm) . . . . . 95
6.6 AQuA PESQ MOS values (Pn = 3.0 dBm) . . . . . 96

6.7 AQuA Voice Quality in Percentage (Pn = -4.5 dBm) . . . . . 97
6.8 AQuA Voice Quality in Percentage (Pn = -3.0 dBm) . . . . . 98
6.9 AQuA Voice Quality in Percentage (Pn = -1.5 dBm) . . . . . 98
6.10 AQuA Voice Quality in Percentage (Pn = -0.5 dBm) . . . . . 99
6.11 AQuA Voice Quality in Percentage (Pn = 2.0 dBm) . . . . . 99
6.12 AQuA Voice Quality in Percentage (Pn = 3.0 dBm) . . . . . 100

# List of Tables


3.1 Audio CODECs [1] . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
3.2 MOS ratings reflecting clean transmission media, absent of errors and noise [2] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
3.3 Factors for which PESQ had demonstrated acceptable accuracy [3] . 18
3.4 PESQ is known to provide inaccurate predictions when used in conjunction with these variables, or is otherwise not intended to be used with these variables [3] . . . . . . . . . . . . . . . . . . . . . . . . . 19
3.5 Factors, technologies and applications for which PESQ has not currently been validated [3] . . . . . . . . . . . . . . . . . . . . . . . . 19
3.6 Factors and applications included in the requirement specification and used in the selection phase of the ITU-T P.863 algorithm [4] . . . . . 24
3.7 ITU-T P.863 is not intended to be used with these variables [4] . . . 25
3.8 Test variables for which further investigation is needed, or ITU-T P.863 is subject to claims of providing inaccurate predictions when used in conjunction with these [4] . . . . . . . . . . . . . . . . . . . 25
3.9 Factors, technologies and applications for which ITU-T P.863 has currently not been validated (For further study) [4] . . . . . . . . . . 25
3.10 Comparison features between PESQ, POLQA and AQuA [5] . . . . 27
3.11 AQuA Codec Conformance test [6] . . . . . . . . . . . . . . . . . . . 28
3.12 BER to RxQual conversion . . . . . . . . . . . . . . . . . . . . . . . 32

4.1 AMR codec modes [7] . . . . . . . . . . . . . . . . . . . . . . . . . 46

5.1 R&S CMW software mandatory options for GSM implementation tests [8] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 58
5.2 R&S CMW software mandatory options for WCDMA implementation tests [9] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59
5.3 Phone Under Test general specifications [10] . . . . . . . . . . . . . . 62
5.4 Behringer ULTRA-DI Model DI100 specifications [11] . . . . . . . . 63
5.5 BHLT10 tri-band Power Splitter/Combiner - Characteristics . . . . . 63

5.6 Unallocated (free for use) frequencies . . . . . . . . . . . . . . . . . . . 67
5.7 Expected average MOS values in Clean Speech (3GPP's Technical Report 26.975 section 5) . . . . . . . . . . . . . . . . . . . . . . . . . 71
5.8 ITU-T P.800 Recommendation - Listening Quality Instructions for MOS values . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 75

# Abbreviations and Acronyms

| | |
|---|---|
| **2G** | Second-Generation |
| **3G** | Third-Generation |
| **4G** | Fourth-Generation |
| **3GPP** | 3rd Generation Partnership Project |
| **3SQM** | Single Side Speech Quality Measurement |
| **8PSK** | 8 Phase-Shift Keying |
| **ACELP** | Algebraic Code Excitation - Linear Prediction |
| **AMR** | Adaptive Multi-Rate |
| **AMR-NB** | Adaptive Multi-Rate Narrowband |
| **AMR-WB** | Adaptive Multi-Rate Wideband |
| **AMR-WB+** | Extended Adaptive Multi-Rate Wideband |
| **AQuA** | Audio Quality Analyser |
| **ARQ** | Automatic Repeat Request |
| **BER** | Bit Error Rate |
| **BLER** | Block Error Rate |
| **BSC** | Base Station Controller |
| **BTS** | Base Transceiver Station |
| **C/I** | Carrier-to-Interference Ratio |
| **CDMA** | Code Division Multiple Access |
| **CELP** | Code Excited Linear Prediction |
| **CEPT** | European Conference of Postal and Telecommunications Administrations |
| **CIR** | Carrier-to-Interference Ratio |
| **CMMB** | China Mobile Multimedia Broadcasting |
| **CNG** | Comfort Noise Generator |

| | |
|---|---|
| **CODEC** | Coder/Decoder |
| **CPICH** | Common Pilot Channel |
| **CRC** | Cyclic Redundancy Check |
| **DCS1900** | North American Digital Cellular Service 1900 |
| **DI** | Direct Injection |
| **DL** | Downlink |
| **DSSS** | Direct-Sequence Spread Spectrum |
| **DTX** | Discontinuous Transmission |
| **DVB-T** | Digital Video Broadcasting - Terrestrial |
| **Ec/N$_0$** | Received Energy per Chip to the Power Density Ratio |
| **EDGE** | Enhanced Data rates for Global Evolution |
| **EFR** | Enhanced Full-Rate |
| **EMI** | Electromagnetic Interference |
| **ETSI** | European Telecommunication Standards Institute |
| **FDD** | Frequency Division Duplexing |
| **FEC** | Forward Error Correction |
| **FER** | Frame Erasure Rate |
| **FFT** | Fast Fourier Transform |
| **FHSS** | Frequency-Hopping Spread Spectrum |
| **FM** | Frequency Modulation |
| **FR** | Full-Rate |
| **GMSK** | Gaussian Minimum Shift Keying |
| **GPRS** | General Packet Radio Service |
| **GPS** | Global Positioning System |
| **GSM** | Global System for Mobile communications |
| **HD** | High-Definition |
| **HF** | High Frequency |
| **HR** | Half-Rate |
| **HSDPA** | High-Speed Downlink Packet Access |
| **HSPA** | High-Speed Packet Access |
| **HSUPA** | High-Speed Uplink Packet Access |

| | |
|---|---|
| **IP** | Internet Protocol |
| **IMS** | IP Multimedia Subsystem |
| **IRS** | Intermediate Reference System |
| **ISDN** | Integrated Services Digital Network |
| **ITU** | International Telecommunication Union |
| **ITU-T** | ITU Telecommunication Standardization Sector |
| **KPI** | Key Performance Indicator |
| **LF** | Low Frequency |
| **LPC** | Linear Predictive Coding |
| **LTE** | Long-Term Evolution |
| **MIC** | Microphone |
| **MNB** | Measuring Normalizing Blocks |
| **MOS** | Mean Opinion Score |
| **MOS-LQO** | Mean Opinion Score - Listening Quality Objective |
| **MP3** | MPEG-1/2 Audio Layer 3 |
| **MPEG** | Moving Picture Experts Group |
| **MS** | Mobile Station |
| **MSC** | Mobile Switching Centre |
| **NB** | Narrowband |
| **PAMS** | Perceptual Analysis Measurement System |
| **PC** | Personal Computer |
| **PCL** | Power Control Level |
| **PCM** | Pulse-Code Modulation |
| **PDC** | Personal Digital Cellular |
| **PDCH** | Packet Data Channel |
| **PESQ** | Perceptual Evaluation of Speech Quality |
| **PN** | Pseudo-random Noise |
| **POLQA** | Perceptual Objective Listening Quality Assessment |
| **PRBS** | Pseudo Random Binary Sequence |
| **PSQM** | Perceptual Speech Quality Measure |
| **PSTN** | Public Switched Telephone Network |

| | |
|---|---|
| **QoS** | Quality of Service |
| **QPSK** | Quadrature Phase Shift Keying |
| **R&S** | Rohde & Schwarz |
| **RAN** | Radio Access Network |
| **RBER** | Residual Block Error Rate |
| **RF** | Radio Frequency |
| **RMC** | Reference Measurement Channels |
| **RMS** | Root Mean Square |
| **RPE** | Regular Pulse Excitation |
| **RPE-LTP** | Regular Pulse Excitation - Long Term prediction |
| **RSCP** | Received Signal Code Power |
| **RSS** | Radio SubSystem |
| **RSSI** | Received Signal Strength Indicator |
| **Rx** | Receiver |
| **RxLev** | Receiver Level |
| **RxQual** | Receiver Quality |
| **SID** | Silence Descriptor |
| **SIM** | Subscriber Identity Module |
| **SNR** | Signal-to-Noise Ratio |
| **SQI** | Speech Quality Indicator |
| **SWB** | Super Wideband |
| **T-DMB** | Terrestrial - Digital Multimedia Broadcasting |
| **TCH** | Traffic Channel |
| **TCX** | Transform Coded Excitation |
| **TD-SCDMA** | Time Division Synchronous Code Division Multiple Access |
| **TDD** | Time Division Duplexing |
| **TDMA** | Time Division Multiple Access |
| **TFO** | Tandem Free Operation |
| **TR** | Technical Report |
| **Tx** | Transmitter |
| **UE** | User Equipment |

| | |
|---|---|
| **UHF** | Ultra High Frequency |
| **UL** | Uplink |
| **UMTS** | Universal Mobile Telecommunication System |
| **UTRA** | UMTS Terrestrial Radio Access |
| **VAD** | Voice Activity Detection |
| **VHF** | Very High Frequency |
| **VoIP** | Voice over Internet Protocol |
| **VoLTE** | Voice over Long-Term Evolution |
| **VSELP** | Vector-Sum Excited Linear Prediction |
| **VSWR** | Voltage Standing Wave Ratio |
| **WB** | Wideband |
| **WCDMA** | Wideband Code Division Multiple Access |
| **WLAN** | Wireless Local Area Network |

# 1 Introduction

The wireless cellular industry is in constant development and it has shown an exponential growth in the number of mobile subscribers since the market appearance of 2G technologies. However, in the last few years the cellular market has become saturated, mostly due to the fact that a very significant part of world's population is already a costumer. By the time of writing the number of mobile subscriptions is roughly around 5760 million [12].
Therefore, operators try to reduce costs, maximize the usage of resources and maintain high quality not only to attract new customers but also to hold current ones.

The key to a satisfied customer is the Quality of Service (QoS). Service quality can be defined as "*the collective effect of service performances which determine the degree of satisfaction of a user of the service*" [13]. In other words, quality is the customer's perception of a delivered service. Hence, management of service quality plays a crucial role if a network operator wishes to thrive in such a competitive industry.
Management of service-quality can be referred to as the monitoring and maintenance of end-to-end services. Services like the one in emphasis in this work, the speech quality as perceived by the end user, are of major importance. Many factors contribute to the degradation of speech quality, e.g., signal quality and noise, making the impact of network performance on the quality of service quite complex, being this matter what motivates this thesis.

The following sections are meant to facilitate the comprehension of the context, goals, schedule and structure of what is presented in this document.

## 1.1 Context

This project emerges from the Thesis/Dissertation subject, of the Master in Electrical and Computer Engineering - Telecommunications Specialization course.
Performed in Óbudai Egyetem (Óbuda University), Kandó Kálmán Faculty of Electrical Engineering - Institute of Communication Engineering. The Thesis main purpose is to contribute to the consolidation of previously acquired knowledge in the specialization field of Telecommunications and, at the same time, to promote the acquiring of new competences.

This work comprises the use of an intrusive method for assessing the speech quality of a radio channel, with experimental and investigative purposes. More specifically aimed at the behavioural comprehension of a set of 2G and 3G speech codecs, correlation between those different technologies and specific transmission parameters.
From a motivational point of view the choice for this project was based primarily on the interest of deepening the knowledge in the field of mobile communications while learning something new in particular subjects, such as:

- Speech Codecs;
- Telephonometry;
- Methods for assessment of speech quality;
- Transmission parameters as Speech quality indicators (SQIs).

## 1.2 Objectives

The main goal of this thesis work is to evaluate the speech quality performance of similar bit rate codecs between different cellular technologies, in clean speech and errors conditions. Followed by further study on how specific transmission parameters can behave as speech quality indicators. This is to be performed with a Wideband Radio Communication Tester that simulates all the protocol layers needed for the wireless cell in test.
Defined objectives are broadly presented below:

- Study of the Wideband Radio Communication Tester;
- Study of methods for evaluation of speech quality;
- Study of speech transmission process, more specifically the source coding - codecs, involved (for 2G and 3G);
- Study of transmission parameters;
- Design of a system architecture for the experiment which fulfils the objectives;
- Evaluation of the results.

## 1.3 Development Plan

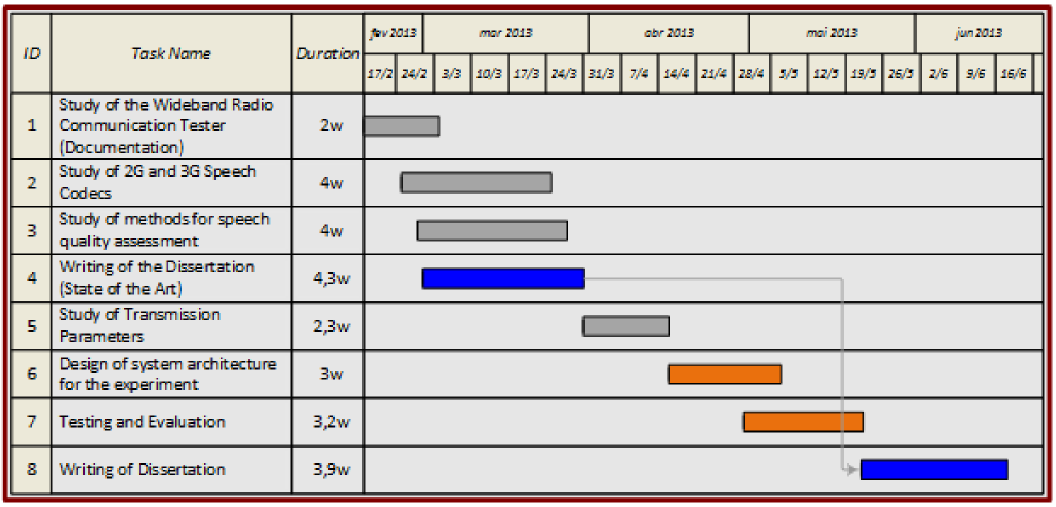



**Figure 1.1:** *Gantt chart regarding the development plan for the Thesis/Dissertation*

A Gantt chart with the development plan for the Thesis project can be consulted in Figure 1.1. Tasks have been separated into three colours, representing, accordingly:

- Gray: Consolidation of the knowledge obtained during research through deeper study of particularly pertinent aspects for the development stage;
- Orange: Development stage including the design of the system architecture, tests and interpretation of results;
- Blue: Writing of the dissertation document.

## 1.4 Outline of the Thesis

In this first chapter the context and contents of the document are briefly described along with the organization and proposed goals.

In the second chapter a general introduction on the 2G and 3G technologies is made, where some important historical, technical and statistical information is presented.

The third and fourth chapters address the theoretical foundations that are recognized as essential to the understanding of some of the topics covered, as well as to the development of this work.

The fifth chapter starts by further contextualizing the project and objectives and, presenting the essential requirements and solution for achieving what is proposed. Then on a second phase of this chapter, all the implementation process is presented along with its tests and analysis of results.

The sixth and final chapter presents the overall achievements, limitations found along the way, final results and conclusion. The chapter is concluded by referring what could be improved or added in future studies.

# 2

# 2G and 3G General Overview

The wireless industry has grown dramatically in the recent years. This is chiefly due to the advances introduced in second-generation Global System for Mobile communications (GSM) technologies and its later acceptance as a universal mobile communications technology. Its development started in 1982, by the predecessor of European Telecommunication Standards Institute (ETSI), the European Conference of Postal and Telecommunications Administrations (CEPT), having defined the following goals:

- Design a superior and more efficient technical solution for wireless communications - it had become obvious at that time that digital systems would be superior in respect to user capacity, ease of use, and number of possible additional services compared with the then-prevalent analogue systems;

- Creating a single standard to be used all over Europe, enabling roaming across borders, as analogue systems incompatibility between countries was a big limitation until then.

The success of GSM exceeded all expectations, and though it was originally developed to be used in Europe, soon was recognized as a worldwide accepted system. Today it is the most popular cellular technology, and by the third trimester of 2010 it had already a user base of over 4000 million in more than 219 countries and territories worldwide [12]. GSM has a wide spectral flexibility - 450, 850, 900, 1800 and 1900 MHz bands (operating with 200 kHz wide channels). So, due to GSM phones

with tri-band and quad-band capability being quite common nowadays, it is rare to find users with no area coverage when travelling to different regions of the globe.
GSM uses Time Division Multiple Access (TDMA) technology and is the legacy network which provides foundation for the third-generation technologies, the Universal Mobile Telecommunication System (UMTS) (also sometimes referred by its main technology's name, Wideband Code Division Multiple Access - WCDMA) and High-Speed Packet Access (HSPA) that we so well know today.
Although originally being a circuit-switched network ideal for the delivery of voice, GSM was designed to evolve. In 2000, the introduction of General Packet Radio Service (GPRS) added packet-switched functionality to its network architecture, thus boosting the delivery of the internet on mobile handsets. Latest enhancements for GPRS were presented in Releases R'98 and R'99, introducing theoretical values for the downlink speed of up to 171 kbit/s.
The next advance in GSM radio access technology was EDGE (Enhanced Data rates for Global Evolution) or Enhanced GRPS. With a new modulation technique yielding a three-fold increase in bit rate (8 Phase-Shift Keying (8PSK) replacing the original digital modulation used, the Gaussian Minimum Shift Keying (GMSK)) and new channel coding for spectral efficiency, EDGE was successfully introduced without disrupting the frequency re-use plans of existing GSM deployments. The increase in data speeds up to 384 kbit/s placed EDGE as an early pre-taste of 3G.
The latest relevant GSM data evolutions were introduced in 3rd Generation Partnership Project (3GPP) Release 7, with the Evolved EDGE, or EDGE II. This release enhances GSM data performance even more, through application of various new techniques.

The success of second-generation cellular telephony motivated the development of a successor system. In between all the advances made in GSM, true 3G technologies stopped being just an idea, but actually became feasible. Mostly as a result of the developments in the GSM Radio Access Network (RAN) Architecture, as aforementioned. The UMTS is a voice and high-speed data technology that uses WCDMA as its radio multiple access technique. UMTS builds on GSM and its main benefits include high spectral efficiency for voice and data, simultaneous voice and data for users, high user densities supportable with low infrastructure costs, high-bandwidth data applications support and migration path to Voice over Internet Protocol (VoIP) [14].
As it is for GSM, UMTS can also work over a wide range of spectrum bands - 850, 900, 1700, 1800, 1900, 2100 and 2600 MHz (operating with 5 MHz wide channels).

However, as opposed to its predecessor, new modes of transmission were introduced, the Frequency Division Duplex (FDD) and Time Division Duplex (TDD). Enabling, in FDD, different frequencies for downlink (DL) and uplink (UL) transmissions, being this the basic system used. As for TDD, this allows the DL and UL to share the same spectrum. UMTS FDD is designed to operate in the paired bands shown in 3GPP Technical Specification (TS) 25.101, section 5.2.a, and their Transmitter-Receiver frequency separation can be consulted in the same Technical Specification, section 5.3.a. UMTS TDD as it does not require frequency separation, instead each TDMA frame gets its timeslots allocated to either transmit or receive depending on the need. Frequency bands designed to operate in UMTS TDD can be consulted in 3GPP TS 25.102.
Subsequent improvements were included in later releases, in particular HSPA, which was first designed for the downlink (High-Speed Downlink Packet Access - HSDPA) and later, also for the uplink (High-Speed Uplink Packet Access - HSUPA). It is basically an extension/improvement of the performance of existing WCDMA protocols. HSPA improves the end-user experience by increasing peak data rates.

Fourth-generation technologies are just now beginning to take place in the commercial plane of mobile communications, but already is seen as a long-term solution for the years to come. Similar to what happened with 2G, in a way, but on the other hand already shows something that 3G did not, an overall acceptance as a universal standard. This being owed mostly to the fact that is an all-IP (packet-switched) solution.
An exponential evolution in mobile communications technologies is expected, as large increasing number of ideas for new developments and improvements are on the run. Also, and as mentioned before, by the time of writing the number of mobile subscriptions is roughly around 5760 million, and it is estimated that 3G subscriptions already surpass GSM [12].

# 3 Speech Quality

Speech quality in wireless cellular networks is affected by a number of factors, such as signal quality, transmission rate, as well as the presence of noise. Noise impairs the conversational quality both for the person placing a call, as well as for the person on the receiving end. During a call one might also experience impairments introduced by the handset or its surroundings. So, the monitoring of speech quality is of great interest to network operators and becomes essential not only to keep the satisfaction of the network users but also to engage new subscribers.

Speech quality can be monitored by means of intrusive and non-intrusive methods. The first is subject of focus in this paper since it meets the requirements intended for the Project.
Intrusive measures incorporate knowledge of the human perceptual system, therefore delivering the nearest results to the ones of subjective tests, like the ITU Telecommunication Standardization Sector (ITU-T) P.800 MOS which will be introduced further along in this chapter. Intrusive measures are based on a comparison between the original speech sample and the transferred (degraded) sample. These are based on algorithms that use psychoacoustic models of human perception, attempting to provide an objective mathematical description of the human perception of sound and, in the process, trying to find variables that have a direct impact on the perceived quality of the voice signal. Intrusive methods comprise several standardized algorithms, like Perceptual Evaluation of Speech Quality (PESQ) and Perceptual Objective Listening Quality Assessment (POLQA), both of which are characterized

in Chapter 3.2.
Non-intrusive methods are another type of objective measurement. These do not need the use of a reference signal and, the final speech quality is calculated only from the parameters of the transferred sample. Despite enabling live network monitoring and assessment by use of unknown speech sources at the far-end side of a telephone connection, such methods are less accurate and less reliable than intrusive ones. An example of a non-intrusive method is ITU-T P.563 Single Side Speech Quality Measurement (3SQM).
In the context of this work, it seems important to understand the basics of Speech Transmission and what makes current wireless cellular networks viable in the first place. In this line of thought, a light background on the matter is given, starting with the principles of a speech transmission system as well of the source coding algorithms that are the base for current speech codecs. Then, a description is made, of the approaches utilized for end-to-end speech quality assessment, from the ITU-T P.800 standard subjective Mean Opinion Score metric, to some of its equivalent objective methods.
Lastly, transmission parameters that help to evaluate the performance of a communication channel, and that in some way, indirectly aid in the assessment of speech quality, are presented.

## 3.1 Speech Transmission

The typical goals of transmission system design such as high capacity, but also high signal quality and robustness, usually contradict each other. Thanks to Shannon's landmark contribution to information theory [Shannon 1948] we are able to treat the issues of bit rate reduction (source coding) and error protection (channel coding) separately. Figure 3.1 illustrates how in general, for mobile communications, the speech transmission system works.
To transmit speech via the physical GSM/UMTS channel, the speech signals have to first be translated into digital signals. This process should maintain a certain speech quality while keeping the required data rate as low as possible, source coding is therefore essential in this matter. Source encoding can follow quite different coding philosophies such as waveform coding, Model-Based coding (also known as parametric coding or vocoding), or hybrid coding, the latter being the most attractive for use in the mobile technologies such as GSM and UMTS.
Waveform coding tries to encode the waveform itself in an efficient way. These codecs are designed to map the input waveform of the encoder in a way that makes it

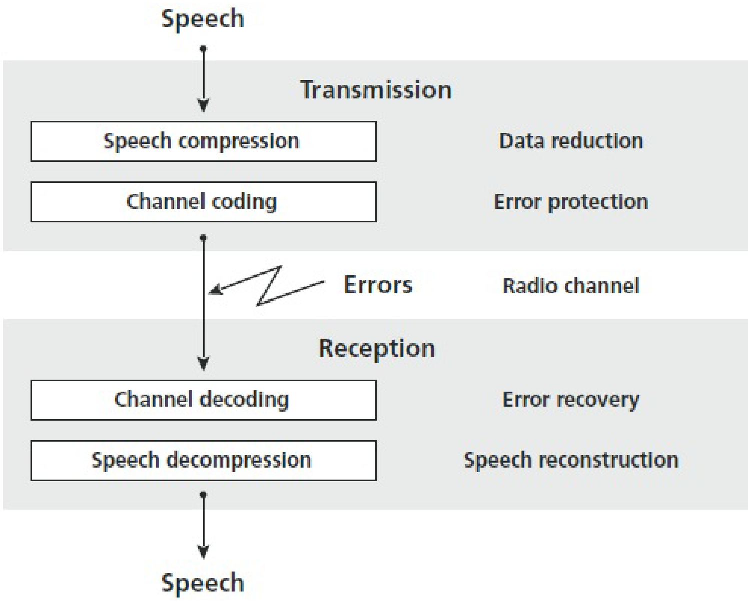



**Figure 3.1:** *General Speech Transmission system process*

possible for the decoder to extract an almost exact replica of the original signal. The decoder essentially inverts encoder processing to restore a faithful approximation of the original waveform. As a result of this advantageous property this type of coding can not only be applied to speech but also to other types of information such as signalling tones, voice-band data, or even music. Naturally, because of this transparency, the coding efficiency (compression) is usually quite modest, though it can be improved by exploiting some statistical signal properties, if the codec parameters are optimized for the most likely categories of input signals, while still maintaining good quality for other types of signals.

The simplest form of this coding type is Pulse-code modulation (PCM) encoding the signal. In general, a signal can be processed further to reduce the amount of storage needed for the waveform, though typically such techniques show some losses: the decoded data differs from the original data. Furthermore waveform codecs can be subdivided into time-domain and frequency-domain codecs.

Waveform coding methods simply try to model the waveform as closely as possible, but it is possible to explore the fact that we are using speech information, to greatly reduce the required storage space. Vocoding techniques do this by encoding information about how the speech signal was produced by the human vocal system, rather than encoding the waveform itself. It uses a small set of parameters which the encoder estimates, quantizes, and transmits over the digital channel. The decoder uses the received parameters to control a real-time implementation of the source model (model of the vocal system) that generates the decoded speech signal. Only recently, vocoders have advanced to a level where we are allowed to use very low rate applications (2.4 kbit/s and below) with few quality constraints [15].

Hybrid coding methods combine the best characteristics between waveform coding and vocoding, both in terms of speech quality and transmission bitrate, although generally at the price of higher complexity (Figure 3.2).

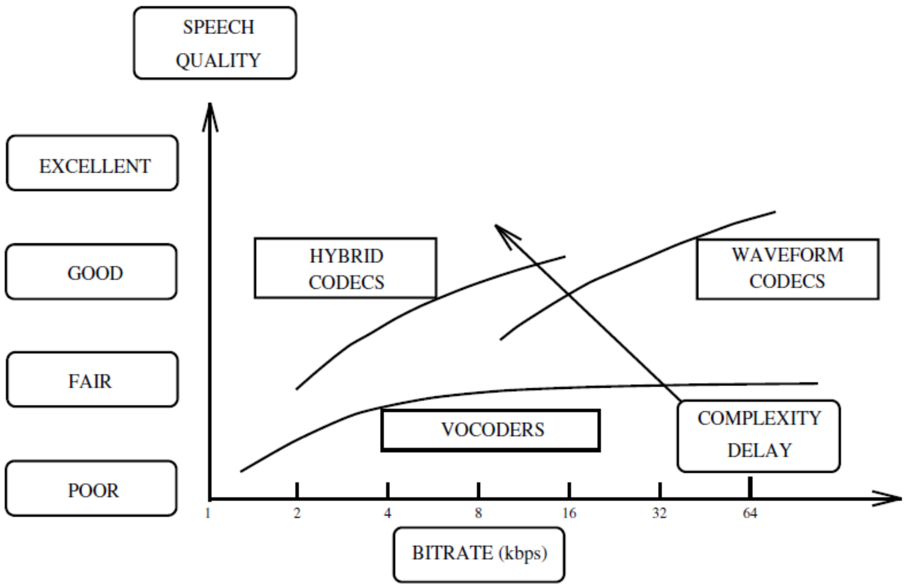



**Figure 3.2:** *Speech quality versus bitrate classification of speech codecs [16]*

Hybrid coders utilize a Model-Based approach to extract speech signal parameters but still compute the modelling error explicitly on the waveform level. This model error or residual waveform is transmitted using a waveform coder whereas the model parameters are quantized and transmitted as side information. The two information streams are combined in the decoder to reconstruct a faithful approximation of the waveform such that hybrid coders share the asymptotically lossless coding property with waveform coders. Their advantage lies in the explicit parameterization of the speech model which allows us to exploit more advanced models than is the case with pure waveform coders which rely on a single invertible dynamical system for their design [15].
The hybrid coders such as Regular Pulse Excitation (RPE) and Code Excited Linear Prediction (CELP) classes are the ones employed in GSM and UMTS CODECs, and will be contextualized in Chapter 4.

Aside from source coding, channel coding is also an essential part of the speech transmission process. Channel coding scheme is generally a combination of error detection, error correcting, rate matching and interleaving[1]. Transmissions over wireless channels are subject to errors, for example due to variations in the received signal quality. To some degree, such variations can be counteracted through link adaptation, which is a mechanism matching automatically transmission parameters (power and data rate) to the channel. However, receiver noise and unpredictable interference variations are inextricably linked to these transmissions and cannot be counteracted. Therefore, virtually all wireless communications systems employ some form of error correction. Error correction exists in two main forms: Automatic Repeat Request (ARQ) and Forward Error Correction (FEC). With ARQ the receiver requests retransmission of data packets, if errors are detected, using some error detection mechanism. In FEC some redundancy bits are added to the data bits, which is done either with parity bits, block wise (so called block coding), convolutional or via Turbo Codes. Convolutional codes provide higher coding gain for small information blocks, e.g. in control signalling, whereas Turbo code is best for larger information blocks of high data rates. As a norm GSM employs convolutional codes, and UMTS convolutional and Turbo Codes.

From an engineering point of view, understanding how the speech transmission process works is only a fundamental part in order to be able to analyse the behaviour of a voice call. Starting with the codecs and type of Error Correction used, to the modulation scheme, these are topics that should be assimilated beforehand.

## 3.2 Codecs and Speech Quality Assessment

A CODEC is a term used to refer to a COder/DECoder and defines a given compression/decompression algorithm or technique that is typically used to convert analogue information (such as speech) into a digital stream for transmission, and then back to analogue at the receiving end. The compression techniques used for voice data and music or other audio data usually differ from each other. As mentioned in the previous section, the explanation for this comes from the simple fact that it is possible to explore certain human voice characteristics in order to further reduce the bandwidth. Then again, as these speech codecs are designed for a speech signal,

[1]The basic premise of interleaving is to spread error bursts over many words of information such that each received word only exhibits at most a few simultaneous symbol errors, which can be corrected for.

they will not reproduce music as well because they will throw away parts of the original signal not expected to be there.
Table 3.1 shows a list of examples of common audio codecs that are currently in use.

**Table 3.1:** *Audio CODECs [1]*

| Standard | Bit rate (kbit/s) | Delay (ms) | MOS | Sample size |
|---|---|---|---|---|
| G.711 | 64 | 0.125 | 4.3 | 8 |
| GSM-FR | 13 | 20 | 3.7 | 260 |
| G723.1 | 6.3/5.3 | 37.5 | 3.9/3.62 | 236/200 |
| UMTS AMR | 12.2-4.75 | Variable | Variable | Variable |
| MP3 | Variable | Variable | Variable | Variable |

Some of these are used in mobile communication networks, for instance GSM, while others are recommended for use with UMTS and IP (G723.1 and G.711, the latter also known as PCM). From the table, MP3 is the only codec not optimized for speech, being normally used on the internet for music coding.
For a voice codec to be chosen, first a set of characteristics have to be taken into account. Ideally the least possible bandwidth is to be used but this generally comes at the expense of quality. Correspondingly, the need for an international standard rating became vital.

### 3.2.1 Subjective Evaluation - MOS

In response, the International Telecommunication Union (ITU) published in August 1996 the P.800 standard recommendation promoting a Mean Opinion Score (MOS) metric, employing five values representing the perceived speech quality of a phone call:

- Excellent = 5
- Good = 4
- Fair = 3
- Poor = 2
- Bad = 1

The MOS for a given codec is relatively subjective, as it is calculated by requesting a number of volunteers to listen to speech and score each sample appropriately. The data is then processed statistically in order to obtain a mean value - the MOS. Since test conditions and human perceptions vary a bit, different studies produce diverse scores, but on the whole, conclusions have been fairly consistent, as reflected by the unofficial ratings given in Table 3.2.

**Table 3.2:** *MOS ratings reflecting clean transmission media, absent of errors and noise [2]*

| Quality of Speech | Score |
|---|---|
| Toll-grade | 4-4.6 |
| Mobile-to-PSTN | 3.5-4.2 depending on the specific vocoder employed |
| Mobile-to-mobile | 3.1-4.1 depending on the specific codec pair and ordinary measurement variations |

When conducting subjective evaluations of speech quality, the standards call for the following requirements that need to be satisfied for the test to yield proper results [2][17]:

- listening is done in a quiet room with a controlled noise level;
- subjects listen through a telephone handset with a standard response;
- speech material should consist of simple, meaningful, short sentences, chosen at random and easy to understand (from current non-technical literature or newspapers, for example). These sentences should aim at fitting into a time-slot of 2-3 seconds, and a minimum of two and a maximum of five sentences are recommended. Typically the standard time length of recordings is of 8 to 12 seconds long (approximately);
- experiments are performed with speech from several different talkers (typically two male, two female) for each coding condition, in order to provide a balance factor. However, since sophisticated processes often affect male and female voices differently, scores should be evaluated separately, only to be averaged if they yield main effects and interactions that are not statistically different;
- subjects are non-expert, which means they are chosen at random from the normal telephone using population, with the provisos that: they have not been directly involved in work connected with assessment of the performance of telephone circuits, or related work such as speech coding. Moreover, they must have not participated in any kind of subjective test for at least the previous six months, and not in a conversation test for at least one year.

After completion of the test, the votes are summed up and a mean opinion score (MOS) is calculated. The corresponding MOS gives the quality of each condition.

However, studying speech quality using subjective scoring was found to be both expensive and time consuming, revealing the need for an inexpensive laboratory tool that could emulate human perception of speech quality in order to assist in the design, testing, and fine-tuning of new codecs[2][17]. Objective methods, as those of the following sections, were then created.

### 3.2.2 Perceptual Evaluation of Speech Quality (PESQ)

PESQ emerged in the year 2000 following the realization that correlation between an earlier release in 1996, the P.861 Perceptual Speech Quality Measurement (PSQM) and subjective opinion scores were poor.

PSQM measure was designed to compute MOS-LQO (where LQO stands for "Listening Quality Objective") values, and similarly to PESQ its main purpose was the assessment of speech codecs, being the first to evaluate and score speech on a quality scale. In 1998, another procedure, Measuring Normalizing Blocks (MNB), was proposed as an optional appendix to P.861 to solve some of PSQM shortcomings. However, neither P.861 PSQM nor MNB were found to be suitable for end-to-end measurement of speech quality in networks. Neither PSQM nor MNB were able to handle variable delay or background noise conditions.

In 2001, ITU-T approved PESQ as recommendation P.862. PESQ combined the best parts of Perceptual Analysis Measurement System (PAMS[2]) and PSQM to yield a significantly higher correlation with subjective studies than any of its predecessors [2].

The PESQ tool was designed to calculate MOS-LQO values on speech samples consistent with subjective evaluations. To do this, PESQ extracts a score from the difference between a reference signal and an output signal emerging from equipment in the signal path. Generally, the greater the difference between the reference signal and the output signal, the lower the MOS value.

Figure 3.3 presents the PESQ architecture. The procedure starts by aligning both reference and degraded signals to a pre-determined loudness level. Following the level adjustment, the signals are then processed by a Fast Fourier Transform (FFT) with an input filter that emulates distortions brought about by processing in a telephone handset. The two signals, the reference and the degraded signals, are then aligned in time and are processed through an auditory transform, a psycho-acoustic model which maps the signals into a representation of perceived loudness in

[2]PAMS is a methodology for assessing speech quality MOS in which its key enhancement over PSQM focuses on end-to-end measurements utilizing time and level alignment and equalization respectively.

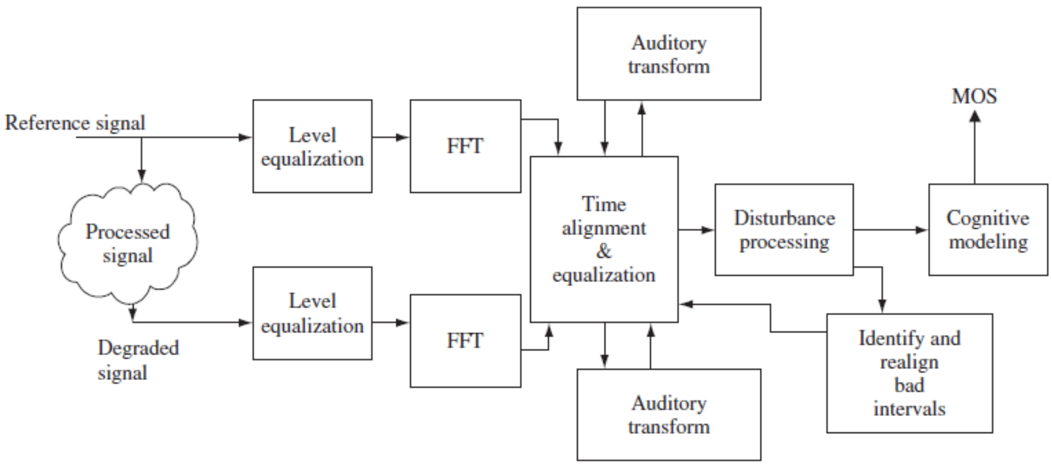



**Figure 3.3:** *PESQ Architecture [2]*

time and frequency, similar to that of PSQM, including equalization of the signals for the frequency response of the system and for gain variation. The disturbance, which is the absolute difference or the gap between the reference and the degraded transforms, is processed in frequency and time. When the first time alignment yields large errors with incorrect delay, the signals are realigned and the disturbance is re-calculated. The process is repeated until it yields a lower disturbance value, and only then is it moved forward to the next step of aggregation where short intervals and the whole signal are considered [2].

Based on the benchmark results presented within 'Study Group 12'[3], an overview of the test factors, coding technologies and applications to which the P.862 PESQ Recommendation applies is given in Tables 3.3 to 3.5. Table 3.3 presents the relationships of test factors, coding technologies and applications for which this Recommendation has been found to show acceptable accuracy. Table 3.4 presents a list of conditions for which the Recommendation is known to provide inaccurate predictions or is otherwise not intended to be used. Finally, Table 3.5 lists factors, technologies and applications for which PESQ has not currently been validated. Although correlations between objective and subjective scores in the benchmark were around 0.935 for both known and unknown data, the PESQ algorithm cannot be used to replace subjective testing. It should also be noted that the PESQ algorithm does not provide a comprehensive evaluation of transmission quality. It only meas-

[3]Study Group, in this context, is a technical group constituted by ITU-T membership agents which take care of the standardization work and develop Recommendations (standards) for the various fields of international telecommunications.

ures the effects of one-way speech distortion and noise on speech quality. The effects of loudness loss, delay, side tone, echo, and other impairments related to two-way interaction (e.g. Centre clipper) are not reflected in the PESQ scores. Therefore, it is possible to have high PESQ scores, yet poor quality of the connection overall [3].

**Table 3.3:** *Factors for which PESQ had demonstrated acceptable accuracy [3]*

| **Test factors** |
|---|
| Speech input levels to a codec |
| Transmission channel errors |
| Packet loss and packet loss concealment with CELP codecs |
| Bit rates if a codec has more than one bit-rate mode |
| Transcodings |
| Environmental noise at the sending side |
| Effect of varying delay in listening only tests |
| Short-term time warping of audio signal |
| Long-term time warping of audio signal |
| **Coding technologies** |
| Waveform codecs, e.g. G.711, G.726, G.727 |
| CELP and hybrid codecs $\geqslant$ 4kbit/s, e.g. G.728, G.729, G.723.1 |
| Other codecs: GSM-FR, GSM-HR, GSM-EFR, GSM-AMR, CDMA-EVRC, TDMA-ACELP, TDMA-VSELP, TETRA |
| **Applications** |
| Codec evaluation |
| Codec selection |
| Live network testing using digital or analogue connection to the network |
| Testing of emulated and prototype networks |

It is recommended that PESQ is used only for speech quality assessment of narrowband (300 to 3400 Hz) handset telephony and narrowband speech codecs. So and for this reason there was the need to create a new standard, one that could not only assess the speech quality of the current telephony systems but also a standard that succeeds on the future short-term evolutions [3].

**Table 3.4:** *PESQ is known to provide inaccurate predictions when used in conjunction with these variables, or is otherwise not intended to be used with these variables [3]*

| **Test factors** |
| --- |
| Listening levels |
| Loudness loss |
| Effect of delay in conversational tests |
| Talker echo |
| Sidetone |
| **Coding technologies** |
| EVRC family codecs |
| Replacement of continuous sections of speech making up more than than 25% of active speech by silence (extreme temporal clipping) |
| **Applications** |
| In-service non-intrusive measurement devices |
| Two-way communications performance |

**Table 3.5:** *Factors, technologies and applications for which PESQ has not currently been validated [3]*

| **Test factors** |
| --- |
| Packet loss and packet loss concealment with PCM type codecs |
| Temporal clipping of speech |
| Amplitude clipping of speech |
| Talker dependencies |
| Multiple simultaneous talkers |
| Network information signals as input to a codec |
| Artificial speech signals as input to a codec |
| Music as input to a codec |
| Listener echo |
| Effects/artifacts from operation of echo cancellers |
| Effects/artifacts from noise reduction algorithms |
| Bit-rate mismatching between an encoder and a decoder if a codec has more than one bit-rate mode |
| **Coding technologies** |
| CELP and hybrid codecs <4kbit/s |
| MPEG4 HVXC |
| **Applications** |
| Acoustic terminal/handset testing, e.g. using HATS |

### 3.2.3 Perceptual Objective Listening Quality Analysis (POLQA)

In 2011, ITU-T approved recommendation P.863, also known as POLQA which incorporates current industry requirements as well as networks and codecs that introduce time warping. POLQA represents the eventual replacement of PESQ as the main standard for speech quality measurements. After all, POLQA was designed specifically to fix some of the reigning PESQ algorithm's known weaknesses, such as inaccuracies with wideband scenarios and CDMA codecs, sensitivity to certain GSM/WCDMA network conditions, and VoIP limitations (variable delay larger than one second, and time scaling).
POLQA also shows greater flexibility than PESQ in providing consistent measurements across the three available bandwidths (narrowband, wideband, and super wideband, or NB, WB, and SWB), different technologies such as GSM, WCDMA and Long-Term Evolution (LTE), both circuit-switched and packet-switched (VoIP, and VoIP over IP Multimedia Subsystem (IMS)), and a variety of interface combinations (electrical-electrical, acoustical-acoustical, electrical-acoustical, and acoustical-electrical). Being the speech quality evaluation algorithm designed for 4G networks, POLQA is best positioned to cope with today's and tomorrow's variety of terminals, new speech codecs and sophisticated error concealment schemes, voice enhancement devices, as well as 4G network types of degradations [18].
A representation of the evolution of ITU-T Recommendations for Speech quality Testing (P.86x), Network Technologies and Bands used, is shown in Figure 3.4:

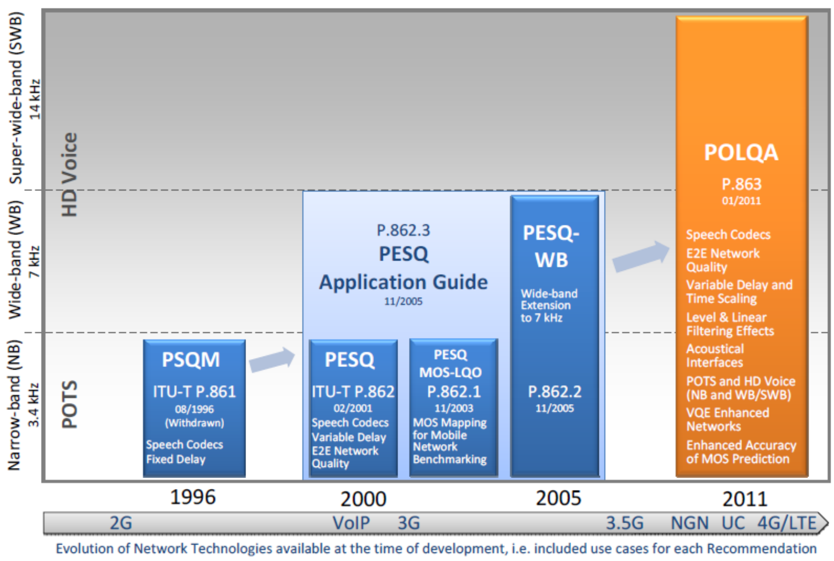



**Figure 3.4:** *Evolution of ITU-T Recommendations for Speech quality Testing (P.86x), Network Technologies and Bands used [19]*

POLQA has two operational modes: NB and SWB. In NB mode, the received (degraded) speech signal is compared with its narrowband (300 to 3400Hz) reference speech. Thus, normal telephone band limitations are not considered to be severe degradations. This narrowband mode offers backward compatibility to P.862.1 (PESQ), which modelled the listening quality as perceived by human listeners, when the degraded speech signal is presented using a loosely coupled Intermediate Reference System (IRS) type handset at one ear (monotic presentation). For a large number of NB scenarios, PESQ and POLQA NB mode showed statistically equal performance [20].

In SWB mode, the received (and potentially degraded or band limited) speech signal is compared with its super wideband reference. The model will output an objective listening quality score as perceived by a human listener using a diffuse-field equalized headphone with diotic presentation (same signal in both ears). Therefore, NB or WB limitations are considered to be degradations and are scored accordingly on a unique SWB MOS scale. In the case of NB quality, it is likely that quality will be compressed at the lower end of the MOS scale, which could impact POLQA SWB performance in predicting NB scenarios. Another reason why NB scenarios in SWB mode are less accurately predicted than in NB mode is the different optimizations that POLQA uses for each of the modes. The NB scenarios are more accurately predicted in the NB mode, since the POLQA optimization is NB focused in this mode [18].

In its main principle the algorithm used in ITU-T P.863 POLQA is similar to the one used in ITU-T P.862 PESQ since it compares a reference signal with a signal which is degraded from passing through a communication system simulating the loss of quality as intended.

The first stage of POLQA processing is temporal alignment of the reference and degraded signal to ensure the following processing in core model is based on an accurate comparison of the same speech segment in two signals. In a first step the reference and degraded signal are split into very small time slices referred as frames. Then the delay of each reference signal frame relative to the associated degraded signal frame is calculated. The sample rate of the degraded signal is then estimated. If sampling rate difference is detected between both signals, the signal with the higher sample rate will be down sampled and the delays re-determined. Once the correct delay is determined and the sample rate differences have been compensated, the signals and the delay information are passed on to the perceptual mode, as shown in Figure 3.5 [21][4].

The key to this process is the transformation of both signals to an internal representation analogous to the psychophysical representation in the human auditory system, taking into account the perceptual pitch (Bark) and the loudness (Sone). This is achieved in several consecutive stages:

- Time alignment;
- Level alignment to a calibrated listening level;
- Time-frequency mapping;
- Frequency warping;
- Compressive loudness scaling.

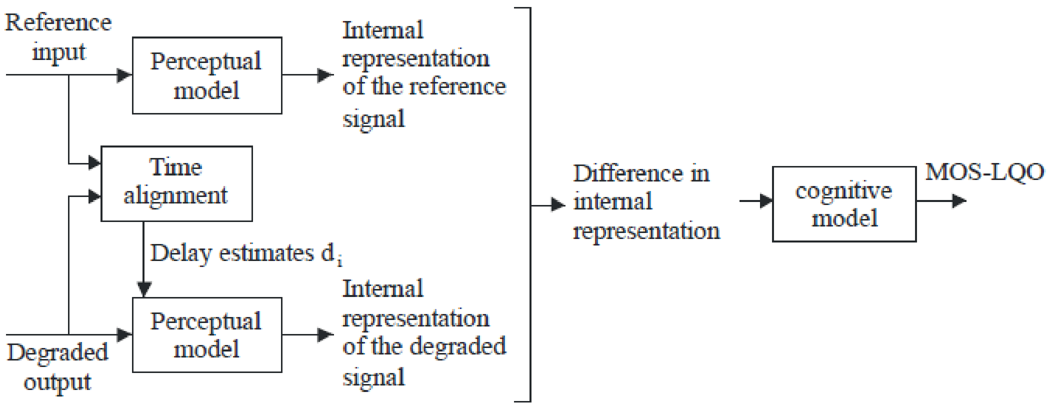



**Figure 3.5:** *Basic POLQA philosophy [4]*

POLQA takes the playback level for the perceived quality prediction into account in SWB mode. In NB mode the speech quality is determined with a constant listening level. By processing the internal representation level, local (rapid) gain variations and linear filtering effects can be taken into account [21].
POLQA also eliminates low noise levels in the reference signal and partially suppresses noise in the degraded output signal. Operations that change the characteristics of the reference and degraded signal are used for the idealization process. This subjective testing is carried out without direct comparison to the reference signal. It supplies six quality indicators that are computed in the cognitive model [21]:

- Frequency response indicator (FREQ);
- Noise indicator (NOISE);
- Room reverberation indicator (REVERB);

- Three indicators describing the internal difference in the time-pitch-loudness domain.

These indicators are combined to give an objective listening quality MOS. POLQA always expects a clean (noise-free) reference signal.

Based on the benchmark results presented within the studies of ITU-T, an overview of the test factors, coding technologies and applications to which this Recommendation applies is given in Tables 3.6 to 3.9. Table 3.6 presents factors and applications included in the requirement specification and which were used in the selection phase of the ITU-T P.863 algorithm. Table 3.7 presents a list of conditions for which this Recommendation is not intended to be used. Table 3.8 presents test variables for which further investigation is needed, or for which ITU-T P.863 is subject to claims of providing inaccurate predictions when used in conjunction with these. Finally, Table 3.9 lists factors, technologies and applications for which the ITU-T P.863 algorithm has not currently been validated [4]. Note that, similarly to PESQ or any other tool (as the one on the next section, AQuA), ITU-T P.863 algorithm cannot be used to replace subjective testing.

Detailed analysis of the PESQ to POLQA transition shows that the new ITU-T P.863 standard is the long term solution for the evaluation of voice service quality as perceived by subscribers. In most cases, operators are expected to move to POLQA once their WB voice service is widely deployed or full Voice over LTE (VoLTE) service solution is implemented in both networks and devices. The expected time frame for this is 2012-2013 [18].

**Table 3.6:** *Factors and applications included in the requirement specification and used in the selection phase of the ITU-T P.863 algorithm [4]*

| **Test factors** |
|---|
| Speech input levels to a codec |
| Transmission channel errors |
| Packet loss and packet loss concealment |
| Bit rates if a codec has more than one bit-rate mode |
| Transcodings |
| Acoustic noise in sending environment |
| Effect of varying delay in listening only tests |
| Short-term time warping of audio signal |
| Long-term time warping of audio signal |
| Listening levels between 53 and 78 dB(A) SPL in superwideband mode |
| Packet loss and packet loss concealment with PCM type codecs |
| Temporal and amplitude clipping of speech |
| Frequency Response |
| Linear distortions, including bandwidth limitations and spectral shaping ('non-flat frequency responses') |
| **Coding technologies** |
| GSM-FR, GSM-HR, GSM EFR |
| AMR-NB, AMR-WB (ITU-T G.722.2), AMR-WB+ |
| PDC-FR, PDC-HR |
| EVRC (ANSI*TIA-127-A), EVRC-B (TIA-718-B) |
| Skype (SILK V3, iLBC, iSAC and ITU-T G.729) |
| Speex, QCELP (TIA-EIA-IS-733), iLBC, CVSD (64 kbit/s, "Bluetooth") |
| MP3, AAC, AAC-LD |
| ITU-T G.711, ITU-T G.711 PLC, ITU-T G.711.1 |
| ITU-T G.718, ITU-T G.719, ITU-T G.722, ITU-T G.722.1, ITU-T G.723.1 |
| ITU-T G.726, ITU-T G.728, ITU-T G.729 |
| **Applications** |
| Codec evaluation |
| Bandwidth extensions |
| Live network testing using digital or analogue connection to the network |
| Testing of emulated and prototype networks |
| UMTS, CDMA, GSM, TETRA, WB-DECT, VoIP, POTS, PSTN, Video Telephony, Bluetooth |
| Voice Activity Enhancement Devices (VED), Noise Reduction (NR) |
| Discontinuous Transmission (DTX), Comfort Noise Insertion |
| Terminal testing, influence of the acoustical path and the transducer in sending and receiving direction. |

**Table 3.7:** *ITU-T P.863 is not intended to be used with these variables [4]*

| **Test factors** |
|---|
| Effect of delay in conversational tests |
| Talker echo |
| Sidetone |
| Acoustic noise in receiving environment |
| **Applications** |
| Non-intrusive measurements |
| Two-way communications performance |

**Table 3.8:** *Test variables for which further investigation is needed, or ITU-T P.863 is subject to claims of providing inaccurate predictions when used in conjunction with these [4]*

| **Test factors** |
|---|
| Acoustical recordings using free-field microphones without HATS or ear-canal simulators |
| **Coding technologies** |
| EVRC family codecs (EVRC, EVRC-B, EVRC-WB, EVRC-NW and others) |

**Table 3.9:** *Factors, technologies and applications for which ITU-T P.863 has currently not been validated (For further study) [4]*

| **Test factors** |
|---|
| Talker dependencies |
| Multiple simultaneous talkers |
| Network information signals as input to a codec |
| Artificial speech signals as input to a codec |
| Music as input to a codec |
| Listener echo |
| Bit-rate mismatching between an encoder and a decoder if a codec has more than one bit-rate mode |
| **Coding technologies** |
| Coding technologies operating below 4 kbit/s |

### 3.2.4 Audio Quality Analyser (AQuA)

Apart from ITU-T standard speech quality assessment most successful tools, PESQ and the most recent POLQA, there are other as efficient tools on the market. Sevana Oy AQuA is an example of that. AQuA provides perceptual estimation and monitoring of audio quality and can be utilized in VoIP, Public Switched Telephone Network (PSTN), Integrated Services Digital Network (ISDN), GSM, WCDMA, LTE/4G networks and combinations of those. This technology allows for evaluation of voice, High-Definition (HD) voice, and wideband audio signals, as it serves for all the above mentioned network types. AQuA is in constant update with the purpose of optimizing the way the algorithm assesses the audio for the evaluation, depending on the various characteristics of the system under test. This way, AQuA is able to adapt to the actual environment in which is being used, rather than being restricted to a standard way of evaluation, thus getting closer to the real perceptual MOS or PESQ MOS values. MOS score values range from 1 to 4.5, considering the assumption that nobody can distinguish between MOS 4.5 and MOS 5.
Analysis of possible reasons for voice and audio quality loss can be made through the analysis results that are stored in a log file, like for example: Signal-to-Noise ratio (SNR) being this the default parameter, and some additional such as delay of audio signal activity, percentage of corrupted signal spectrum, or significant distortion in low, and medium frequency bands.
In a way, AQuA is not only a competitive tool when compared with ITU-T PESQ, but it approaches POLQA in terms of features and capabilities, being able to work with a majority of technologies, such as GSM, WCDMA, LTE, VoIP, VoIP over IMS, some of these mentioned before. Table 3.10 shows some of the other main features in comparison with ITU-T speech quality assessment tools.

In the same way as PESQ and POLQA, AQuA Model (Figure 3.6) compares a reference sample with its correspondent degraded sample.

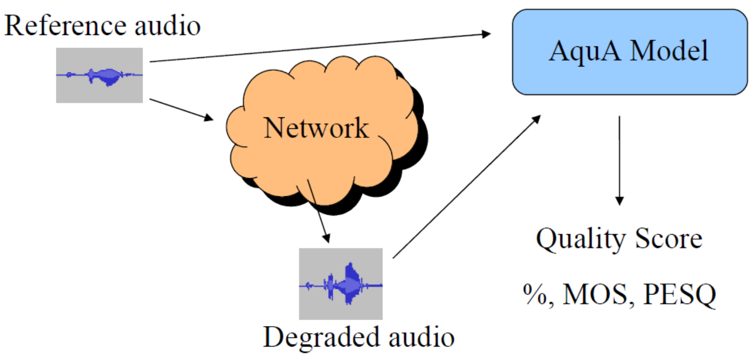



**Figure 3.6:** *AQuA Model [6]*

**Table 3.10:** *Comparison features between PESQ, POLQA and AQuA [5]*

| Feature | PESQ | POLQA | AQuA |
|---|---|---|---|
| VAD detection in AMR codec | ✗ | ✓ | ✓ |
| GSM/WCDMA accurate measurements | ✗ | ✓ | ✓ |
| Variable delay up to 1 second in VoIP | ✗ | ✓ | ✓ |
| Strong linear distortions | ✗ | ✓ | ✓ |
| Works in SWB bandwidths | ✗ | ✓ | ✓ |
| Considers network specific degradations | ✗ | ✓ | ✓ |
| Noisy listening environments | ✗ | ✗ | ✓ |
| Music | ✗ | ✗ | ✓ |
| MOS score for long sequences of speech (more than 12 seconds) | ✗ | ✗ | ✓ |
| Considers codec degradations (tandem codecs, packet loss, frame errors, bit errors) | ✗ | ✓ | ✓ |

The approach done in AQuA's algorithm is quite different from PESQ or POLQA, but the basic and fundamental principle is the same, taking advantage of specific features and characteristics inherent to humans. As mentioned before in Chapter 3.1. and according to the well-known fact, sound signals quality is determined not only by the technical characteristics of a sound processing and transfer systems, but also by the properties of individual peculiarities of speech perception and production, which vary in time and from individual to individual.
What made the conception of the current AQuA algorithm a possibility was the discovery, from different field researchers (Zwiker, Fletcher, Sapozhkov, Pokrovskij), of certain bands in speech signal which influence the quality of perceived speech. As these bands varied so did the precision of the quality score. Sevana Oy joined research of all the involved scientists, merging these bands into a specific set of so called critical bands, thus aiding in the creation of the current AQuA algorithm. The value of spectrum energy in bands can be used for different purposes, one of which is the sound signal quality estimation, although depending on the critical bands used it is possible to increase the accuracy of estimation depending on its purpose (example: speech signal instead of sound signal).

Figure 3.7 represents the AQuA general scheme of the quality estimation system for sound signals.

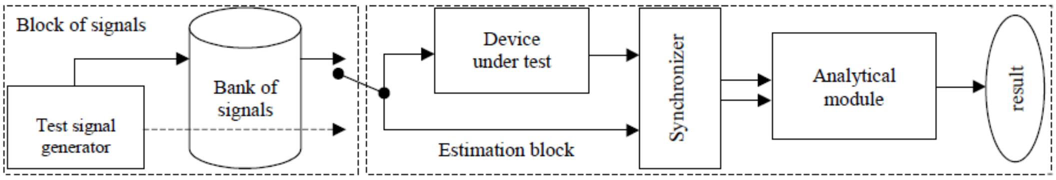



**Figure 3.7:** *AQuA general scheme of the quality estimation system for sound signals [22][23]*

First: "*Generator of test signals allows sound signal forming according to one of the sound flow models. It can be either a particularized set of sound signals or a signal, received in output of statistical speech model. Generator's signal can either be saved for follow-up usage or be exposed to processing and estimation. Bank of signals stores sound data, received as a result of signals' generator work or from some external sources.*" [22][23].
Accordingly, "*an input of estimation block is a signal coming from the generator directly or, one from the bank of signals. The test signal is the input of the synchronizer or of the device under test, which can be for example, a vocoder or a communication channel. The output signal of the device under test is an input of synchronizer also.*"[22][23].
Lastly: "*the synchronizer matches in time an initial signal and a processed signal. The synchronized signals in chunks are then input in the analytical module, which determines the degree of similarity for signals and issues the quality estimation as the measure of similarity between the initial and the processed signals.*"[22][23].
The presented method of sound signal quality estimation shows result estimations well correlated with that of MOS P.800. Table 3.11 illustrates an AQuA codecs conformance test, which corroborates with the above stated.

**Table 3.11:** *AQuA Codec Conformance test [6]*

| Codec | MOS, P.800 | AQuA MOS | ITU PESQ |
|---|---|---|---|
| a-law | 4.1 | 4.18 | 3.0 |
| $\mu$-law | 4.1 | 4.18 | 3.0 |
| g.723.6.3 | 3.9 | 3.9 | 2.93 |
| g.723.5.3 | 3.62 | 3.65 | 2.91 |
| gsm.6.10 | 3.16 | 3.7 | 2.87 |
| g.729 | 3.9 | 3.85 | 4.08 |

To summarize, AQuA and PESQ do the same thing differently, although if one needs to provide end costumers speech quality score according to international standard then PESQ, or its successor POLQA, are the tools to use. Nevertheless, if one needs to assess speech quality with a quite accurate prediction of ITU-T P.800 MOS score, AQuA shows itself as a great solution. Being backed up in this matter by some of its costumers spread out globally: CallFire, Aircell, Ventelo, Mondial Telecom, Binary Elements, Bel Air Internet, Fibernetics, Vodafone, Intermec, Modulis, Celya informatique et communication, Redvoiss, Gentrice, Intersvyaz, Callvine, U-TX, Arkadin, VSB-TUO, CESNET [6].

## 3.3 Transmission Parameters

Monitoring and evaluation in cellular networks is essential to the success of every cellular operator. Key Performance Indicators (KPIs) are defined by the network operator, based on their business needs and these, derived from the behaviour of the network. All in all it is correct to say that the purpose of performance measurement is to troubleshoot and optimize the network.
Optimization involves monitoring, verifying and improving the performance of the radio network. A cellular network covers a large area and provides capacity to many people, so there are lots of parameters involved that are variable and have to be continuously monitored and corrected. As a network is always growing through rising in subscriber numbers and increases in traffic, this means that the optimization process should be on going in order to increase the efficiency of the network and, lead to a revenue generation.

For the purpose of this work, only some KPIs relevant for evaluating speech quality will be focused on in this chapter.
Performance and service quality in GSM networks is usually measured by means of the parameters like Receiver Quality (RxQual), Bit Error Rate (BER), Frame Erasure Rate (FER) (also sometimes referred to as Frame Error Rate) and L3 messages. Though traditionally, RxQual parameter is the most commonly used to measure the quality, however, it suffers from a number of drawbacks which make it an unreliable indicator of speech quality if one means to use this parameter alone. Still, some studies show good correlations between RxQual and speech quality, in which RxQual is demonstrated as a great tool for the network managers (getting them a quite good idea of the impact in the end-to-end speech quality).
Figure 3.8 displays a typical KPI distribution for different terminal models taken from the most commonly used ones in a commercial GSM network.

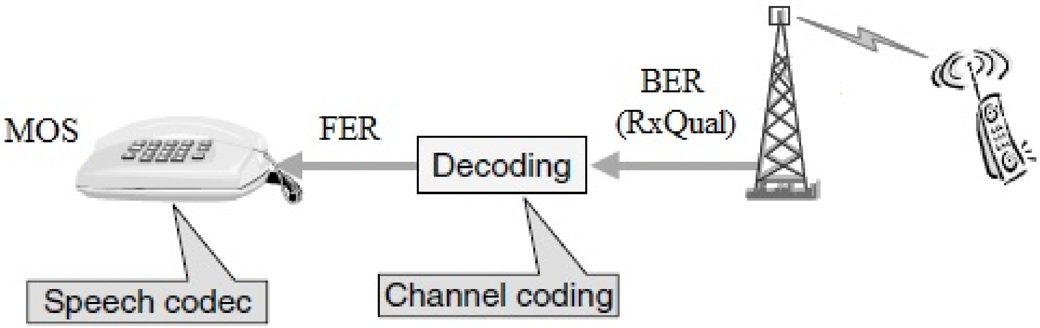



**Figure 3.8:** *Main speech Key Performance Indicators in GSM [24]*

As for the UMTS networks, Received Signal Code Power (RSCP), Ratio of the Received Energy per chip and the Noise Power Density ($Ec/N_0$ ), Carrier-to-Interference Ratio (CIR or C/I), and Block Error Rate (BLER) are the main parameters utilized for evaluation.

### 3.3.1 Bit Error Rate

Noise disrupts the quality of communication between sender and receiver as the received noisy voltage samples can cause the receiver to incorrectly identify the transmitted bit, thereby generating a bit error. If we transmit a long stream of bits and count the fraction of received bits that are in error, we obtain a quantity called the Bit Error Rate. This quantity is equivalent to the probability that any given bit is in error [25].

Bit Error Rate is a key parameter that is used in assessing systems that transmit digital data from one location to another. BER is applicable to radio data links, ethernet, as well as fibre-optic data systems. When data is transmitted over a data link, there is a possibility of errors being introduced into the system. If this is so, the integrity of the system may be compromised. Hence the necessity of assessing the performance of the system, and BER provides an ideal way in which this can be achieved. BER assesses the full end to end performance of a system including the transmitter, receiver and the medium between the two. BER is defined as the rate at which errors occur in a transmission system. In simple form [26]:

$$BER = \frac{\text{number of bits in error}}{\text{total number of bits sent}} \tag{3.1}$$

BER expression is given by Rappaport in [27], as

$$\int_0^\infty P_b(E/r) \times P(r)dr \tag{3.2}$$

where, $P_b(E/r)$ is the conditional error probability, and P(r) is the probability density function of the SNR.
A bit error occurs when an electrical or optical receiver makes an incorrect decision about a bit's logic level. Many factors can contribute to BER, they include for example: signal power, noise, jitter, multipath effects, and Electromagnetic Interference (EMI) from radiated emissions. The presence of noise at the point where a receiver decides the logic level of a bit can cause the receiver to misinterpret that bit. Poor signal-to-noise ratio (SNR) and poor extinction ratio (the average power level of logic-1 bits compared to the average power of logic-0 bits) make it easier for random noise to cause bit errors [28].
Communication links exhibit a wide range of bit error rates. At one end, high-speed (multiple gigabits per second) fibre-optic links implement various mechanisms that reduce the bit error rates to be as low as one part in $10^{12}$ ($1/10^{12}$). This error rate looks exceptionally low, but a link that can send data at 10 gigabits per second with such an error rate will encounter a bit error every 100 seconds of continuous activity, so it does need ways of masking errors that occur [25].
On the other hand, wireless communication links usually have errors anywhere between one part in $10^4$ for a relatively noisy environments, down to one part in $10^7$. Very noisy links can still be useful even if they have bit error rates as high as one part in $10^2$ or $10^3$.
General performance BER requirements for GSM in particular can be consulted in 3GPP TS 45.005 "*Radio transmission and reception*".
Although BER can somehow indicate speech quality, it alone is not enough. The most important key performance indicator used for this assessment is the one mentioned before, the Mean Opinion Score.

### 3.3.2 Block Error Rate

Block error rate (also sometimes referred as Block error ratio) is an analysis of transmission errors on the radio interface. It is characterized in 3GPP Technical Specification 34.121 (F.6.1.1.) as follows: "*A Block Error Ratio is defined as the ratio of the number of erroneous blocks received to the total number of blocks sent. An erroneous block is defined as a Transport Block, the cyclic redundancy check (CRC) of which is wrong*". This can be expressed in the following equation:

$$BLER(\%) = \frac{\sum \text{RLC Transport Blocks with CRC error}}{\sum \text{RLC Transport Blocks}} \times 100 \tag{3.3}$$

To perform a block error ratio measurement, a Radio Bearer Test Mode call should be set. The WCDMA specifications (3GPP TS 34.109) define a Radio Bearer Test Mode as a suite of uplink (UL) and downlink (DL) reference measurement channels (RMCs) to use for UE transmitter and receiver conformance test. The block error ratio measurement can be performed for both Symmetrical RMCs and Asymmetrical RMCs, with a downlink rate of 12.2, 64 or 144 kbit/s the test set sends a block every 20 ms. As for RMCs with a downlink rate of 384 kbit/s, the test set sends a block every 10 ms [29].

### 3.3.3 Receiver Quality

RxQual, also termed RxQual Full, is a value between 0 and 7, where each value corresponds to an estimated number of bit errors in a number of bursts. Each RxQual value corresponds to the estimated bit-error rate according to Table 3.12, which is taken from 3GPP TS 45.008 section 8.2.4:

**Table 3.12:** *BER to RxQual conversion*

| | |
|---|---|
| RxQual 0 | 0% <BER <0.2% |
| RxQual 1 | 0.2% <BER <0.4% |
| RxQual 2 | 0.4% <BER <0.8% |
| RxQual 3 | 0.8% <BER <1.6% |
| RxQual 4 | 1.6% <BER <3.2% |
| RxQual 5 | 3.2% <BER <6.4% |
| RxQual 6 | 6.4% <BER <12.8% |
| RxQual 7 | 12.8% <BER |

The channel BER is averaged over intervals of 480 ms and mapped to the logarithmic RxQual parameter with the eight BER levels from 0 (BER <0.2 %) to 7 (BER >12.8 %) shown in the table [30]. RxQual serves as an estimate of the current channel quality during an active call, and controls the binary error channel in the GSM simulations performed. In the GSM system, values below four are desirable, because at a gross BER of less than 1.6 %, nearly all bit errors within the most important class-I-bits can be corrected by the channel decoder [31].

### 3.3.4 Receiver Level

The received power level at the mobile station (MS) is measured in dBm and mapped linearly to an RxLev index ranging from 0 (<-110 dBm) to 63 (>-48 dBm) in 1 dBm steps. The minimum required value specified in the GSM standard ranges from -104 to -100 dBm (RxLev <6 . . . 10). Measurements are reported every 480 ms.

The received power level describes the radio channel in terms of path loss and slow fading. However, it is not a measure of signal-to-interference ratio, but really an expression of the sum of the desired signal plus interference. A high correlation with the resulting speech quality is therefore only expected for the case when the interference is low and relatively constant, for example, in a GSM system with a large cluster size [31].
RxLev has an important role in GSM networks handover decisions. Although exact handover strategies are determined by the network operator, the behaviour follows the subsequent rules [30]:

- Inter-cell handover - occurs either when the measurements show low RxLev and/or RxQual on the current serving cell and a better RxLev available from a surrounding cell, or when a surrounding cell allows communication with a lower Transmitter (Tx) power level. This typically indicates that the mobile station is on the border of its serving cell area coverage;

- Intra-cell handover - executed from one channel/timeslot in the serving cell to another channel/timeslot in the same cell, when RxQual is low but there is a high RxLev. This indicates a decrease on the quality, due to an increase in the interference. Therefore, this type of handover should provide a different channel with lower levels of interference as a solution.

### 3.3.5 Frame Erasure Rate

The frame erasure rate is a percentile value and indicates how many of the received frames are bad. The channel decoder erases frames when the CRC fails. FER is performed on both speech and signalling frames. When a speech frame is discarded, the system will interpolate and the mobile station is instructed to resend that same frame. In the GSM standards, it is common to see FER limits multiplied by a correctional value $\alpha$. This correction factor allows the system designer to trade-off between FER and Residual BER (RBER). If the system is designed to discard a large number of frames, then the FER will be higher (multiply by $\alpha$) and RBER should be significantly lower (divide by $\alpha$). Values for $\alpha$ range between 1 and 1,6. So FER can be calculated through [32]:

$$BLER(\%) = \frac{\sum \text{no. of blocks with incorrect CRC}}{\sum \text{total no. of blocks}} \times 100 \tag{3.4}$$

For conversational voice, it has been suggested in 3GPP TS 22.105, that acceptable performance is typically obtained with frame erasure rates up to 3 %. However, to maintain this acceptable performance the percentage should not go higher than 4 %.

### 3.3.6 Carrier-to-Interference Ratio

The signal quality of a connection is measured as a function of received useful signal power and interference power received from co-channel cells and is given by the Carrier-to-Interference Ratio (CIR or C/I) [7]:

$$\frac{C}{I} = \frac{\text{Useful signal power}}{\text{Disturbing signal power}} = \frac{\text{Useful signal power}}{\text{Interference power from other cells}} \tag{3.5}$$

To better visualize this, an uplink situation is illustrated in Figure 3.9.

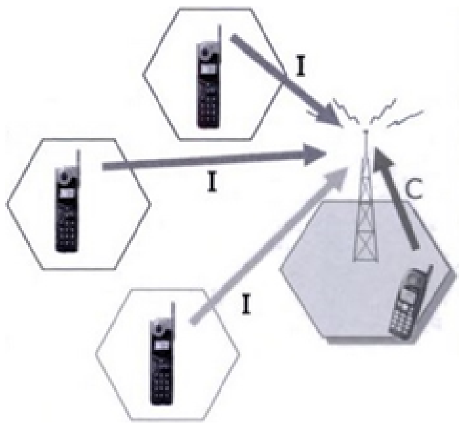



**Figure 3.9:** *Illustration of interference received from co-channels - uplink situation [33]*

The intensity of the interference is essentially a function of co-channel interference depending on the frequency reuse distance D. From the viewpoint of a mobile station, the co-channel interference is caused by base stations at a distance D from the current base station. A worst-case estimate for the C/I of a mobile station, is at the border of the covered cell area at distance R (radius of the cell) from the base station. In this scenario the mobile station is subject to highest propagation losses (C has the lowest value), and it is assumed that the other six neighbouring interfering transmitters operate at the same power and are approximately equally far apart (frequency reuse distance D is much greater when compared with the cell radius R)[7]. For this worst-case scenario CIR can be calculated as:

$$\frac{C}{I} = \frac{1}{M} \times (D/R)^{\gamma} = \frac{1}{6} \times (D/R)^{\gamma} \tag{3.6}$$

, where M is the number of interference sources (for hexagonal cells M=6), and $\gamma$ is the path loss exponent (which depends on the surrounding environment).

WCDMA can work with lower C/I values when compared, for example, with GSM, since in spread spectrum techniques (examples: CDMA, DSSS, FHSS) the wideband signal can be below the thermal noise level. A good quality speech connection in GSM requires C/I 9-12 dB [34][35].

### 3.3.7 Received Energy per Chip to the Power Density Ratio

Generally used ratio in WCDMA networks for the Common Pilot Channel (CPICH), CPICH $Ec/N_0$ is the ratio of the received energy per Pseudo-random Noise (PN)[4] chip of the pilot channel to the total received power spectral density at the UE antenna connector. In other words, the $Ec/N_0$ is the RSCP divided by the Received Signal Strength Indicator (RSSI) (or $RSCP_{dB}$ - $RSSI_{dB}$, in the usual logarithmic scale). The reported range for CPICH Ec/N0 is from -24 dB to 0 dB (0 to 49), and can be consulted in 3GPP TS 25.133 s9.1.2.3.
The higher this value the better can a signal of a cell be distinguished from the overall noise. The $Ec/N_0$ is usually expressed in decibels as it is a relative value. The value is negative as a logarithmic scale is used and the RSCP is smaller than the total received power. It can be used to compare the relative signal quality of different cells on the same carrier (frequency). Their relative difference to each other, independent of their absolute signal strengths, can then be used, for example, to aid on handover or cell reselection decisions [36].

[4] In spread spectrum communications (examples: DSSS, FHSS, CDMA), the term chip or chip sequence refers to a spreading-code sequence also known as pseudo-random noise (PN)

# 4
# Source Coding in GSM and UMTS

Considering the use of different codecs for the implementation of this work, a detailed characterization of speech codecs used in 2G and 3G source coding is made in this chapter.
Starting with GSM and its four standard codecs, Full-Rate (FR), Half-Rate (HR), Enhanced Full-Rate (EFR), and Adaptive Multi-Rate (AMR). Then, UMTS codecs are presented, from the AMR Narrowband (AMR-NB), AMR Wideband (AMR-WB), to the most recent Extended AMR Wideband (AMR-WB+).
To conclude the chapter, a section on Performance Characterization of speech codecs relevant for the work is presented. With the results based on an experiment reported in the Technical Report (TR) 26.975 produced by the 3rd Generation Partnership Project.

## 4.1 GSM Codecs

One of the most important services in GSM is the voice service. Thus, from the start of the service, need for continuous improvements was realised as a priority, not only for the network operators own gain but because the general consumers demanded it. This generalised demand was not pursued directly, at the time of introduction and growth phases of the technology, but rather indirectly since PSTN had obviously the major market share.
The priority was the development of new speech codecs with two competing object-

ives:

- better utilization of the frequency bands assigned to GSM;
- improvement of speech quality to a level that is similar to that offered by ISDN networks, which was the primary request of professional users.

The GSM standard supports four different but similar compression technologies to analyse and compress speech. These include Full-Rate, Half-Rate, Enhanced Full-Rate and Adaptive Multi-Rate. Despite all being lossy (i.e. some data is lost during the compression), these codecs have been optimized to accurately regenerate speech at the output of a wireless link.
For a better understanding on the main characteristics of these codecs, and how these work, the next sub sections present their characterization.

### 4.1.1 Full-Rate codec

When GSM was first being specified, the challenge was to prove that the limited available spectrum could be exploited more efficiently than with the existing analogue systems. That would mean the capacity of systems (i.e., number of customers the mobile network can support for a given amount of licensed frequency allocation) could be maximized whilst preserving, or even improving, the speech quality as perceived by the user. The work resulted in a digital 'full-rate speech' coding algorithm [37].
The coding scheme implemented in FR codec is called Regular Pulse Excitation - Long Term prediction - Linear Predictive Coder, commonly referred to as RPE-LTP. With the GSM full-rate codec, the analogue speech signal is usually sampled with a rate of 8000 samples per second and an 8-bit resolution of the analogue-to-digital conversion, resulting this way in a data rate of 64 kbit/s. The FR speech coder then reduces this rate to 13 kbit/s. Then, to gain robustness against transmission errors (channel coding), these speech-coded data are encoded with a convolutional code, yielding a transmission rate of 22.8 kbit/s [38].

### 4.1.2 Half-Rate codec

The reason for improved bandwidth utilization is to increase the network capacity and the spectral efficiency (i.e., traffic carried per cell area and frequency band), hence decreasing system costs. Therefore, plans to introduce a half-rate speech codec were soon developed and eventually presented with interesting results. Under good channel conditions, this codec achieves, in spite of the half bit rate, a comparable intrinsic quality to the one shown by the full-rate used till then. Though

demonstrating slightly lower speech quality, as it was already expected.
However, quality loss occurs in particular for mobile-to-mobile communication, since in this case (due to the ISDN architecture) one has to go twice through the GSM speech coding/decoding process. These multiple, or tandem, conversions degrade speech quality. The end-to-end transmission of GSM-coded speech is intended to avoid multiple unnecessary transcoding and the resulting quality loss (Figure 4.1). This technique has been passed under the name Tandem Free Operation (TFO) in GSM Release 98 [7].

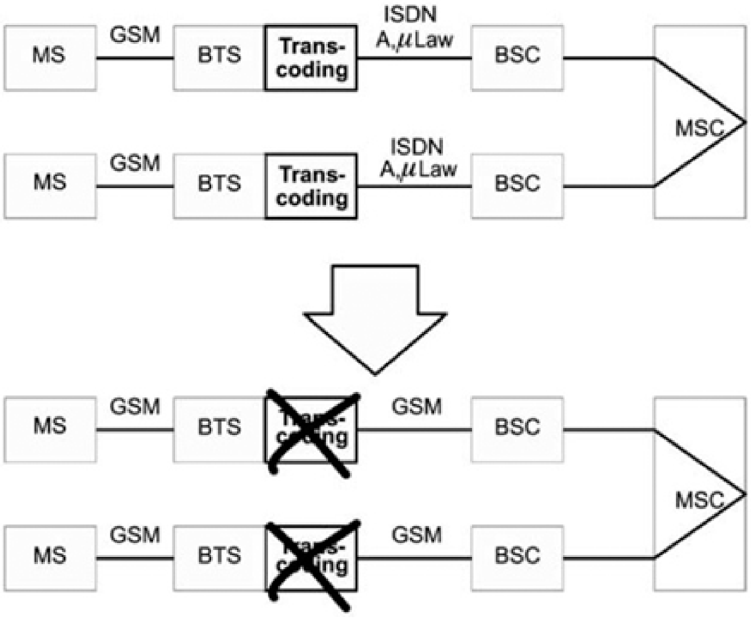



**Figure 4.1:** *Through-transport of GSM-coded speech in Phase 2+ for mobile-to-mobile connections (TFO)*

As referred in 3GPP TS 46.020 section 4, the GSM half-rate codec operates at the rate of 5.6 kbit/s and uses the Vector-Sum Excited Linear Prediction (VSELP) algorithm. The VSELP algorithm is an analysis-by-synthesis coding technique and belongs to the class of speech coding algorithms known as CELP. The GSM half-rate codec's encoding process is performed on a 20 ms speech frame at a time. A speech frame of the sampled speech waveform is read and based on the current and past history of the waveform, the codec encoder extracts 18 parameters that describe it. These parameters are then grouped into the following three general classes [39]:

- energy parameters;
- spectral parameters;
- excitation parameters.

These are quantized into 112 bits for transmission.
The GSM half-rate codec is an analysis-by-synthesis codec, therefore the speech decoder is primarily a subset of the speech encoder. The quantized parameters are decoded and a synthetic excitation is generated using the energy and excitation parameters. The synthetic excitation is then filtered to provide the spectral information resulting in the generation of the synthesized speech (see Figure 4.2) [39].

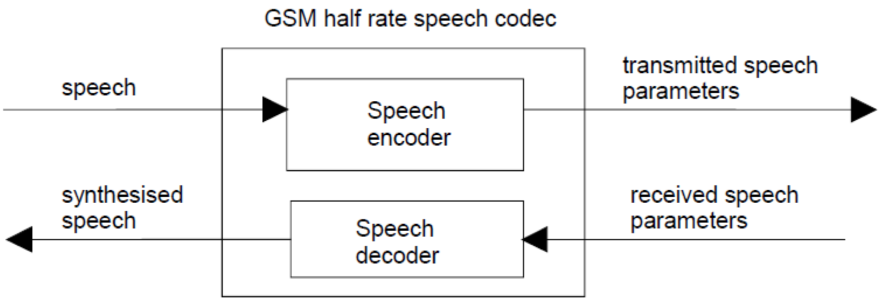



**Figure 4.2:** *Block diagram of the GSM half rate speech codec [39]*

### 4.1.3 Enhanced Full-Rate codec

Reaching a speech quality level that is close to what is found in wireline networks is crucial for business applications and in cases where GSM systems are intended to replace wireline networks, e.g., for fast installation of telecommunication networks in areas with insufficient or missing telephone infrastructure. As a result of the continuous improvement of voice coding techniques it became possible to design vocoders complying with the mentioned demands [7].
Work on the Enhanced Full-Rate codec was therefore considered as high priority. This EFR is a full rate codec (net bit rate 12.2 kbit/s). Nevertheless, it achieves speech quality that is clearly superior when compared to the previously used full-rate codec. It has been initially standardized and implemented in North American Digital Cellular Service 1900 (DCS1900) networks and has been implemented in GSM with high success since it met the requirements set by the European committees. Instead of using the RPE-LTP coding scheme implemented in FR codec, a so-called Algebraic Code Excitation - Linear Prediction (ACELP) is employed [7].
The EFR speech coder delivers data blocks of 244 information bits to the channel encoder. In addition to categorising the bits into important class I bits and less important class II bits, EFR further divides into class Ia bits and class Ib bits. A special preliminary channel coding is employed for the most significant bits: eight parity bits (generated by a CRC) and eight repetition bits are added to provide

additional error detection. The resulting 260 bits are processed by the block encoder [7].

### 4.1.4 Processing functions for FR, HR and EFR

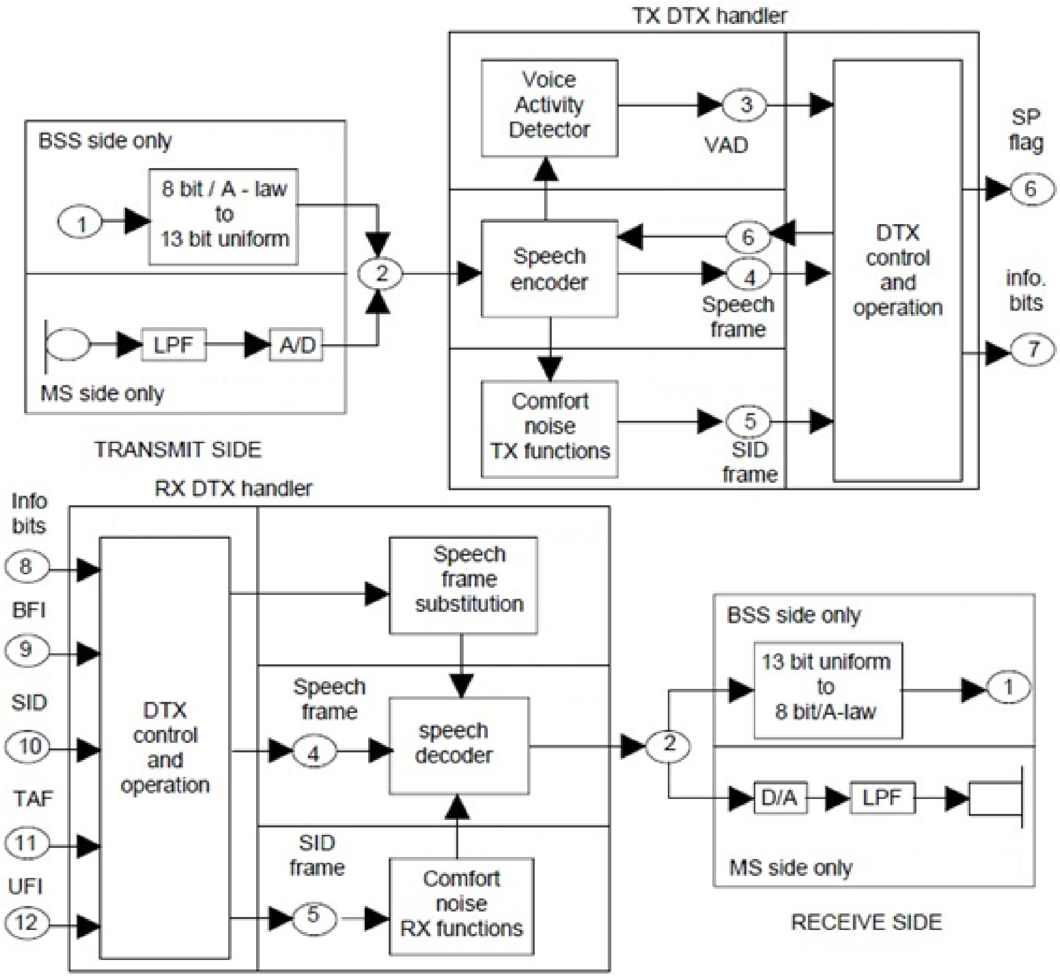
It is relevant to comprehend what are the main processing functions for the above mentioned standard speech codecs, so in this sub section they will be contextualized in the transmission system, as well as characterized. Also, these processing functions are present in most speech codecs in use today (e.g. AMR) for wireless cell communication and, their general concept and purpose are equally applied.
Figure 4.3 presents a reference configuration where the various speech processing functions are identified. The audio parts including analogue to digital and digital to analogue conversion are included to show the complete speech path between the audio input/output in the Mobile Station (MS) and the digital interface of the PSTN [40][41][42].

As shown in Figure 4.3, the speech encoder takes its input as a 13 bit uniform PCM signal either from the audio part of the MS or on the network side, from the PSTN via an 8 bit/A-law or $\mu$-law (PCS 1900) to 13 bit uniform PCM conversion. The encoded speech at the output of the speech encoder is delivered to the channel coding function to produce an encoded block consisting of [40][41][42]:

- 456 bits leading to a gross bit rate of 22.8 kbit/s, in Full Rate and Enhanced Full Rate;
- 228 bits leading to a gross bit rate of 11.4 kbit/s, in Half Rate.

In the Receiver (Rx) direction, the inverse operations take place.
With the sampling rate of 8 000 sample/s, this leads to an average bit rate for the encoded bit stream of [40][41][42]:

- 13 kbit/s (encoded blocks of 260 information bits) in Full Rate. With the 'RPE-LTP - Linear Predictive Coder' coding;
- 5.6 kbit/s (encoded blocks of 112 information bits) in Half Rate. With the VSELP coding scheme;
- 12.2 kbit/s (encoded blocks of 244 information bits) in Enhanced Full Rate. Employing ACELP as the coding scheme.




1 - 8 bit /A-law or $\mu$-law (PCS 1900) PCM (ITU-T Recommendation G.711), 8 000 samples/s.
2 - 13 bit uniform PCM, 8 000 samples/s.
3 - Voice Activity Detector (VAD) flag.
4 - Encoded speech frame, 50 frames/s, 112 bits/frame.
5 - Silence Descriptor (SID) frame, 112 bits/frame.
6* - Speech (SP) flag, indicates whether information bits are speech or SID information.
7 - Information bits delivered to the radio subsystem.
8 - Information bits received from the radio subsystem.
9 - Bad Frame Indication (BFI) flag.
10 - Silence Descriptor (SID) flag.
11 - Time Alignment Flag (TAF), marks the position of the SID frame within the Slow Associated Control Channel (SACCH) multiframe.
12** - Unreliable Frame Indication (UFI).
(* - does not exist in Full Rate)
(** - only exists in Half Rate)

**Figure 4.3:** *Half rate codec - processing functions reference configuration [40][41][42]*

### Discontinuous Transmission (DTX)

During a regular conversation, the participants alternate so that, on the average, each direction of transmission is occupied about 50 % of the time. Discontinuous Transmission (DTX) is a mode of operation where the radio transmitters are switched on only for those frames which contain useful information and, consequently, switched off during speech pauses. 3GPP Technical Specifications [43][44][45] assert that this may be done for the following two purposes:

- to save power in the MS;
- to reduce the overall interference level over the air interface, leading to better Radio Frequency (RF) spectrum efficiency.

The overall DTX mechanism is implemented in the DTX handlers (Tx and Rx sides), instructed by the network and the following functions are required [43][44][45]:

- a Voice Activity Detector (VAD) on the Tx side;
- evaluation of the background acoustic noise on the Tx side, in order to transmit characteristic parameters to the Rx side;
- on the Rx side a similar noise denominated Comfort Noise is generated, during periods where the radio transmission is switched off.

The transmission of comfort noise information to the RX side is accomplished by sending a Silence Descriptor (SID) frame at the end of speech bursts (also serving as an end of speech marker for the Rx side). These SID frames are sent at regular intervals, also during speech pauses, with the purpose of updating the comfort noise characteristics at the Rx side. This also serves for the measurement improvement of the radio link quality by the Radio SubSystem (RSS) [40][41][42].
For the overall DTX functionality, the DTX handlers interwork with the RSS using flags. The RSS is controlled by the transmitter keying on the Tx side, which performs pre-processing functions on the Rx side [40][41][42]. This is described in GSM TS 06.81.
The speech flag (SP) indicates whether information bits are speech or SID information and is calculated from the VAD flag by the Tx DTX handler. When SID information is transmitted (SP="0") the operation of the speech encoder is modified to reduce the remaining computation for that frame [40][41][42]. This is described in GSM TS 06.62.

### Voice Activity Detection (VAD)

The function of the VAD is to distinguish between noise with speech present and, noise without speech present. The principles found in 3GPP TS [46][47][48] describe that the VAD is basically an energy detector that compares the energy of a filtered version of the input signal with a threshold. The presence of speech is indicated whenever the threshold is exceeded.
The detection of speech in mobile environments is not an easy task as a result of the low speech/noise ratios which are encountered, particularly in moving vehicles. It has been found that the noise is relatively stationary for quite long periods in a mobile environment, it is therefore possible to use an adaptive filter with coefficients obtained during noise, to remove much of the vehicle noise. To increase the probability of detecting speech, the input signal is adaptively filtered to reduce its noise content before the voice activity decision is made [46][47][48].
The frequency spectrum and level of the noise may vary within a given environment as well as between different environments. It is therefore necessary to adapt the input filter coefficients and energy threshold at regular intervals [46][47][48].

### Comfort Noise Generator (CNG)

Background acoustic noise constitutes a problem when DTX function is being utilized. This noise is transmitted together with the speech, and would disappear when the radio transmission is switched off, resulting in a modulation of the background noise. Since the DTX switching can take place rapidly, it has been found that this effect may be annoying for the listener, especially, for example, in a car environment with high background noise levels [49][50][51].
Comfort noise was the solution found to overcome this problem, by generating on the Rx side, synthetic noise similar to the Tx side background noise. The comfort noise parameters are estimated on the Tx side and transmitted to the Rx side before the radio transmission is switched off and, at a regular low rate afterwards. This allows the comfort noise to adapt to the changes of the noise on the Tx side[49][50][51].

Finally and in addition to the main processing functions above characterized there is still one of relevant aspect which deserves to me mentioned, one which masks the effect of an isolated lost frame. In this case, the lost speech frame is substituted by a predicted frame based on previous frames (insertion of silence frames is not allowed). In case of several subsequent lost frames, a muting technique is used to indicate to the listener that transmission has been interrupted [40][41][42]. This technique has been proved to be efficient and successfully implemented.

### 4.1.5 Adaptive Multi-Rate codec

An intrinsic property of radio frequency channels, which is a subject of constant study and research, is the fact that the conditions are not always the same. Hence, there is a need to adapt the transmission properties in response to those channel condition changes.

The speech codecs characterized in the previous sub chapters, full-rate, half-rate and EFR, are of fixed source/information bit rate, and have been optimized for typical radio channel conditions. Therefore, if having in mind the above stated, it is easy to understand that this fixed rate approach lacks flexibility: whenever the channel conditions are much worse than the typical, speech quality will suffer significant degradation, since the channel capacity assigned to the MS is too small for error free transmission. Also, radio resources will be wasted for unneeded error protection if the radio conditions are better than usual [34].

To overcome these problems, in October 1997, a new programme was initiated in ETSI, aimed at the development of Adaptive Multi-Rate (AMR) codec for GSM system. The main target of the work was to develop a codec that would provide a significant improvement in error robustness, as well as capacity, over EFR. The actual AMR codec standardization was carried out as a competitive selection process consisting of several phases. In February 1999, ETSI approved the AMR codec standard, which was based on the codec developed in collaboration between Ericsson, Nokia and Siemens. Two months later, 3GPP adopted the AMR codec as the mandatory speech codec for the 3G WCDMA system. Some parts of the AMR codec development, such as voice activity detection (VAD) and optimized channel coding were finalized and included in the standard later, in June 1999 [24].

AMR codec achieves an improved speech quality by adaptively switching between different speech coding schemes (with different levels of error protection) according to the current channel quality. To be more precise, AMR has two principles of adaptability [52]: channel mode adaptation and codec mode adaptation.

Channel mode adaptation dynamically selects the type of traffic channel that a connection should be assigned to: either a full rate (TCH/F) or a half rate traffic channel (TCH/H). The basic idea here is to adapt a user's gross bit rate in order to optimize the use of radio resources. If the traffic load in a cell is high, those connections using a TCH/F (gross bit rate 22.8 kbit/s) and having high channel quality should be switched to a TCH/H (11.4 kbit/s). On the other hand, if the load is low, the speech quality of several TCH/H connections can be improved by

switching them to a TCH/F. The signalling information for this type of adaptation is done with existing protocols on GSM signalling channels (the switching between full-rate and half-rate channels is executed by an intra-cell handover) [34][7].
The task of codec mode adaptation is to adapt the coding rate (i.e. the trade-off between the level of error protection versus the source bit rate) according to the current channel conditions. Whenever the radio channel conditions are poor, the encoder is set to operate at low source bit rates as its input and, to use additional bits for forward error protection. When the quality of the channel is good, less error protection is used [7].
The AMR speech codec utilizes the ACELP algorithm, the same as the one employed in EFR, and combines eight different modes with source/information bit rates ranging from 12.2 to 4.75 kbit/s (see Table 4.1). From the results of link quality measures, an adaptation unit selects the most appropriate codec mode. Figure 4.4 illustrates the AMR encoding principle (with bit numbers for TCH/F - full rate example) [7].

**Table 4.1:** *AMR codec modes [7]*

| | **Source data rate (kbit/s)** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 12.2 | 10.2 | 7.95 | 7.4 | 6.7 | 5.9 | 5.15 | 4.75 |
| Information bits per block | 244 | 204 | 159 | 148 | 134 | 118 | 103 | 95 |
| Class Ia bits (CRC protected) | 81 | 65 | 75 | 61 | 55 | 55 | 49 | 39 |
| Class Ib bits (not CRC protected) | 163 | 139 | 84 | 87 | 79 | 63 | 54 | 56 |
| Rate R of convolutional encoder | 1/2 | 1/3 | 1/3 | 1/3 | 1/4 | 1/4 | 1/5 | 1/5 |
| Output bits from convolutional encoder | 508 | 642 | 513 | 472 | 576 | 520 | 565 | 535 |
| Punctured bits | 60 | 194 | 65 | 26 | 128 | 72 | 117 | 87 |

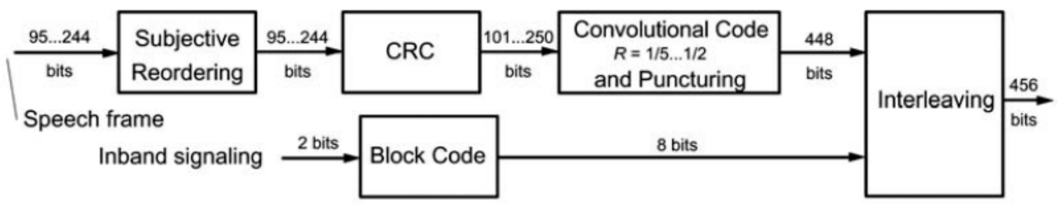



**Figure 4.4:** *AMR channel encoding principle (bit numbers for TCH/F) [7]*

As it is asserted in [7], channel coding is performed using a punctured recursive systematic convolutional code. Since not all bits of the voice data are equally important for audibility, AMR also employs an Unequal Error Protection structure. The most important bits (class Ia; e.g. mode bits and Linear Predictive Coding (LPC) coefficients) are additionally protected by a CRC code with six parity bits.

On the Rx, the decoder will dispose of the entire speech frame if the parity check fails. Also, the degree of puncturing depends on the importance of the bits. A block with a fixed number of gross bits is expected at the end of the encoding process, which is then interleaved with the goal of reducing the number of burst errors [7].
It is certain that channel conditions vary quite rapidly, thus codec mode adaptation requires a fast signalling mechanism. This is achieved by transmitting the information about the used codec mode, link control, DTX, among others, together with the speech data in the TCH, i.e., a special in-band signalling is employed [7].
The example for full rate mode in Figure 4.4 is given: the 12.2 kbit/s codec for a TCH/F operates with 244 source bits (12.2 kbit/s $\times$ 20 ms), which are first reorganised by order of subjective importance. Then six CRC bits for class Ia bits are added, obtaining 250 bits. The subsequent block, the recursive convolutional encoder, is defined by the two generators which maps those bits to 508 bits. Next, 60 bits are punctured, which results in an output sequence of 448 bits. Together with the encoded in-band signalling (8 bits) this block is interleaved and finally mapped to bursts. The resulting gross bit rate is thus 456 bits/20 ms = 22.8 kbit/s [7].

## 4.2 UMTS Codecs

Following the great success on the development of Adaptive Multi-Rate speech codec for GSM, which had been validated after an extensive feasibility study, right away 3GPP felt the need to adopt the AMR codec as its mandatory speech codec for the 3G WCDMA system.
As the needs for continued enhancement and better utilization of the mobile network resources grew, concurrent with the objective of reaching a level of speech quality that surpassed PSTN's quality, new codecs were created to overcome new defined goals.
In the next sub sections the speech codecs in use on 3G technologies are described. From the standard AMR speech codec (later renamed to AMR-NB), to the improved: AMR-WB and AMR-WB+ audio codec.

### 4.2.1 AMR-NB and AMR-WB

The standard speech codec in UMTS employs the AMR technique utilized in GSM. As it is in 2G, here the speech coder is also a single integrated speech codec with eight source rates: 12.2 (GSM-EFR), 10.2, 7.95, 7.40 (IS-641), 6.70 (PDC-EFR), 5.90, 5.15 and 4.75 kbit/s [53]. To ease interoperability with existing cellular networks, some of the codec modes are the same as those already present in those networks: the 12.2 kbit/s AMR speech codec is identical to the GSM EFR codec, 7.4 kbit/s is equal to the North American US-TDMA speech codec, and 6.7 kbit/s is identical to the Japanese Personal Digital Cellular (PDC) codec. Switching between different bit rates is possible on a time interval of each speech frame (20 ms) (in-band signalling is used for this mode switching) [35].

The codec was later changed in name, to AMR narrowband (AMR-NB) since its wideband version was established.

The AMR-NB coder operates on speech frames of 20 ms corresponding to 160 samples at the sampling frequency of 8000 samples per second, whereas AMR wideband is based on the 16 000 Hz sampling frequency, thus extending the audio bandwidth to 50-7000 Hz. The coding algorithm for the multi-rate coding modes is, as mentioned before, ACELP, but is also sometimes referred to as MR-ACELP for the narrow band, and MRWB-ACELP for the wideband codec version. In every 160 speech samples, analysis are made in order to extract CELP model parameters such as, 'LP filter coefficients'[1], adaptive[2] and fixed codebooks'[3] indices and gains [54][55].

The speech parameter bits delivered by the speech encoder are rearranged according to their subjective importance before they are sent to the network. The importance classes are Class A, B and C, where A contains the bits most sensitive to errors and so the strongest channel coding is here applied [56][57].

Basic functions described earlier as, DTX, VAD and comfort noise, are all effectively implemented in AMR. This codec also comprises error concealment, also termed frame substitution and muting procedure. Its purpose is the same as the one used in GSM codecs, which is to conceal the effect of lost AMR speech frames. As mentioned before, the purpose of muting the output in the case of several lost frames is to indicate the interruption of the channel to the user and, to avoid generating possibly annoying sounds as a result of the frame substitution procedure [53][56].

Moreover, the bit rate of the AMR speech connection can be controlled by the Radio Access Network depending on the air interface loading and the quality of the speech connections. During high loading, such as during busy hours, it is possible to use lower AMR bit rates to offer higher capacity while providing slightly lower

[1]Linear Prediction (LP) coefficients (also referred as Linear Predictive Coding coefficients) is a generic descriptive term for the short term filter coefficients)

[2]Adaptive codebook: contains excitation vectors that are adapted for every sub-frame. The adaptive codebook is derived from the long-term filter state. The lag value can be viewed as an index into the adaptive codebook.

[3]Fixed codebook: The fixed codebook contains excitation vectors for speech synthesis filters. The contents of the codebook are non-adaptive (i.e., fixed). In the adaptive multi-rate codec, the fixed codebook is implemented using an algebraic codebook.

speech quality. Also, for cases when the User Equipment (UE) is on the border of its current cell coverage area and using its maximum transmission power, then a lower AMR bit rate can be used to extend the cell coverage area. The AMR speech codec makes it possible to achieve a compromise between the network's capacity, coverage and speech quality according to the operator's requirements [35].

### 4.2.2 Adaptive Multi-Rate-Wideband codec

AMR-WB speech codec was first presented in 3GPP's Release 5 and, introduced considerable voice quality improvements compared with the AMR-NB codec or compared with a standard fixed telephone line. In the case of packet-switched streaming, AMR-WB is already part of Release 4 [35].
In ITU-T Recommendation G.722.2, the AMR-WB is selected as a standard wideband codec (for 16 kbit/s). This is of significant importance, since this is the first time that the same codec has been adopted for wireless and wireline services. This will eliminate the need of transcoding and ease the implementation of wideband voice applications and services across a wide range of communications systems [35][58]. The AMR-WB codec operates on nine speech-coding bit-rates between 6.60 kbit/s and 23.85 kbit/s [59].

The term wideband originates from the sampling rate, which has been increased from 8 kHz to 16 kHz. This allows one to cover twice the audio bandwidth compared with the classical telephone voice bandwidth of 4 kHz. While all the previous codecs in mobile communication operate on a narrow audio bandwidth limited to 200-3400 Hz, AMR-WB extends the audio bandwidth to 50-7000 Hz. Figure 4.5 shows the listening test result, where AMR-WB is compared with AMR-NB. The results are presented as the Mean Opinion Score (MOS). The MOS results show that AMR-WB is able to improve the voice quality without increasing the required radio bandwidth. For example, AMR-WB 12.65 kbit/s provides a clearly higher MOS than AMR-NB 12.2 kbit/s. The improved voice quality can be obtained because of higher sampling frequency [35].

For another set of results that corroborate what is illustrated in Figure 4.5, the reader can consult the tests done in Multimedia Technologies laboratory - Nokia Research Center [60]. Also for an extensive variety of experimental performance characterization tests using AMR-WB, one can consult the 3GPP TR 26.976.

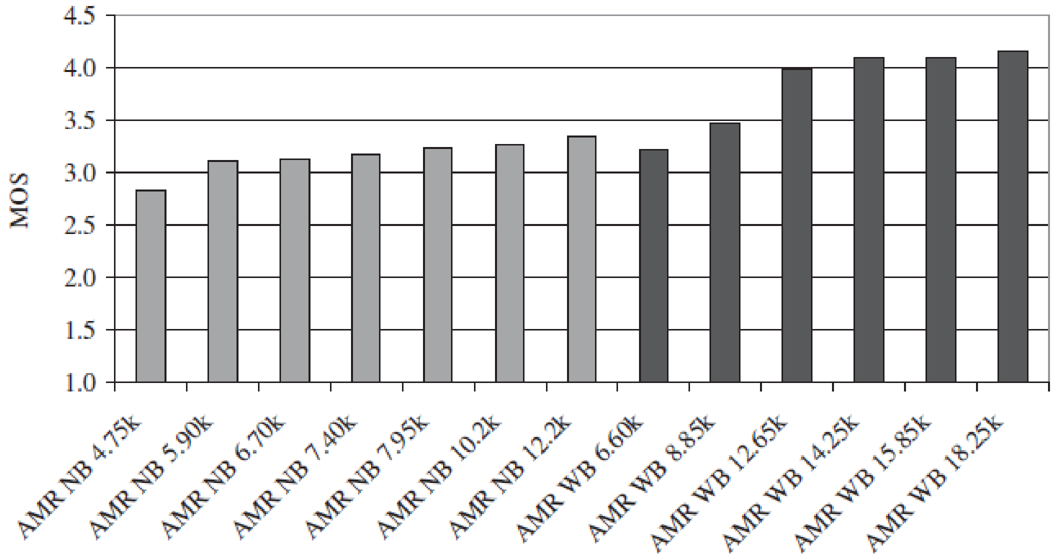



**Figure 4.5:** *Mean opinion score (MOS) example with wideband and narrowband AMR*

### 4.2.3 Extended Adaptive Multi-Rate Wideband codec

Although the current most used speech codecs on UMTS are the ones above mentioned, seems relevant to introduce the latest addition to this class of voice codecs, the AMR-WB+.

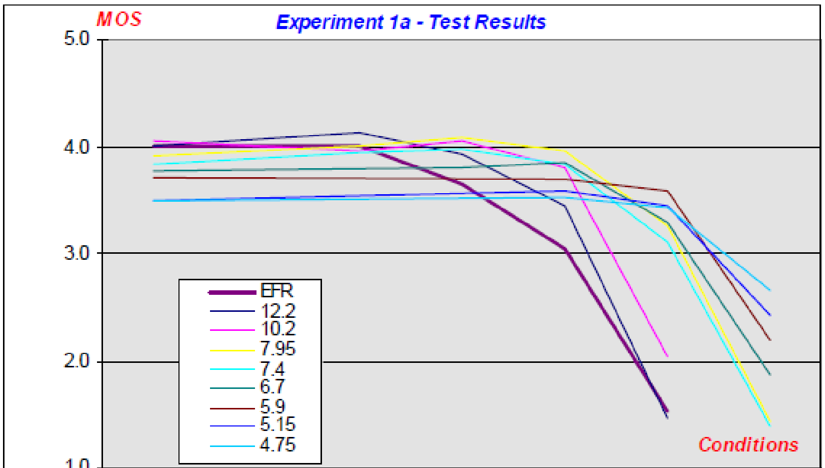
This audio codec comprises all AMR-WB speech codec modes, AMR-WB DTX, VAD and comfort noise, as well as the extended functionality added by transform coded excitation (TCX), bandwidth extension and stereo. These new extended features serve for encoding general audio signals such as music, speech, mixed and other signals.

The encoding algorithm at the core of the AMR-WB+ codec is based on a hybrid ACELP/TCX model. For every block of input signal, the encoder decides (either in open-loop or closed-loop) which encoding model (ACELP or TCX) is best. The ACELP model is a time-domain, predictive encoder, best suited for speech and transient signals. The AMR-WB encoder is used in ACELP modes. Alternatively, the TCX model is a transform-based encoder, and is more appropriate for typical music samples [61].

According to 3GPP TS 26.290 the AMR-WB+ audio codec processes input frames equal to 2048 samples at an internal sampling frequency Fs. The internal sampling frequency is limited to the range 12800-38400 Hz. The 2048-sample frames are split into two critically sampled equal frequency bands. This results in two superframes of 1024 samples corresponding to the low frequency (LF) and high frequency (HF) bands. Each superframe is divided into four 256- samples frames. Sampling at the internal sampling rate is obtained by using a variable sampling conversion scheme, which re-samples the input signal. The LF and HF signals are then encoded using two different approaches. The LF is encoded and decoded using the "core" encoder/decoder, based on switched ACELP and TCX. The HF signal is encoded with relatively few bits (16 bits/frame) using a 'bandwidth extension' method [61].

## 4.3 Performance Characterization of Speech Quality

With the aim of further comparison and interpretation of this work's results, codec performances demonstrated in 3GPP's TR 26.975 section 5 will be taken into account. Section 5 of 3GPP TR 26.975 features two experiments, 1a (Full Rate) and 1b (Half Rate) and, assesses the speech quality (MOS) under clean speech and error conditions. Carrier-to-Interference Ratio is the measured parameter utilized to define a static dB level for each error condition test.




| | No Errors | C/I=16 dB | C/I=13 dB | C/I=10 dB | C/I= 7 dB | C/I= 4 dB | C/I= 1 dB |
|---|---|---|---|---|---|---|---|
| EFR | 4.01 | | 4.01 | 3.65 | 3.05 | 1.53 | |
| 12.2 | 4.01 | 4.06 | 4.13 | 3.93 | 3.44 | 1.46 | |
| 10.2 | 4.06 | | 3.96 | 4.05 | 3.80 | 2.04 | |
| 7.95 | 3.91 | | 4.01 | 4.08 | 3.96 | 3.26 | 1.43 |
| 7.4 | 3.83 | | 3.94 | 3.98 | 3.84 | 3.11 | 1.39 |
| 6.7 | 3.77 | | | 3.80 | 3.86 | 3.29 | 1.87 |
| 5.9 | 3.72 | | | | 3.69 | 3.59 | 2.20 |
| 5.15 | 3.50 | | | | 3.58 | 3.44 | 2.43 |
| 4.75 | 3.50 | | | | 3.52 | 3.43 | 2.66 |

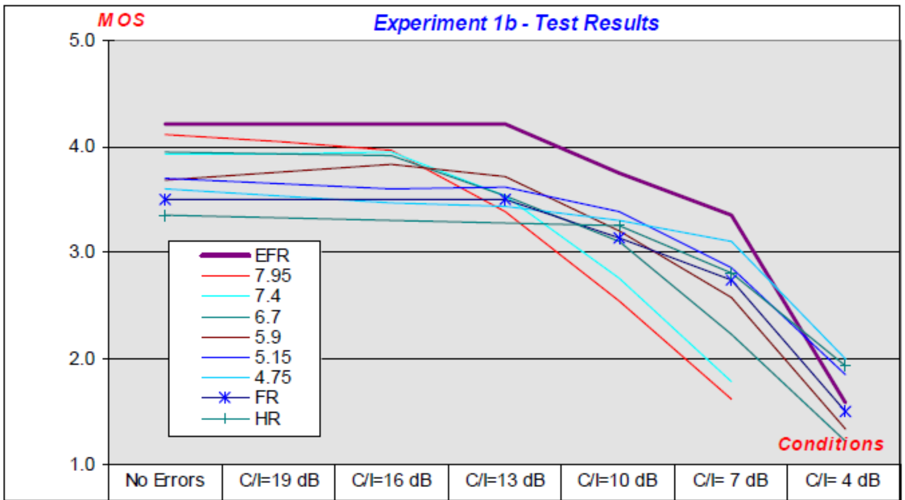
**Figure 4.6:** *Family of curves for Experiment 1a (Clean Speech in Full Rate)*

The tests include the following codecs: GSM EFR, GSM FR, GSM HR, AMR FR and AMR HR.
Figure 4.6 provides a graphical representation (in Mean Opinion Scores) of all 8 AMR full rate codec modes performances, with the corresponding performance of GSM EFR.
For Experiment 1a the test results display that AMR FR speech codec modes provide robust speech quality down to 4 dB C/I. The results also show the four highest codec modes (12.2, 10.2, 7.95 and 7.4) being equivalent to EFR in error free conditions and, barely affected by propagation errors over a wide range channel conditions (down to 10-7 C/I). The four lowest codec modes (6.7, 5.9, 5.15 and 4.75) are all judged in error free conditions to be equivalent to EFR at 10 dB C/I. The three lowest codec modes show the best resilience, not being statistically affected by propagation errors down to 4 dB C/I.

Figure 4.7 provides a graphical representation (in Mean Opinion Scores) of all 6 AMR half rate codec modes for each impairment condition, with the corresponding performance of the GSM EFR, GSM FR and GSM HR speech codecs.

For Experiment 1b, test results demonstrate that all AMR HR speech codec modes provide robust speech quality down to 16 dB C/I. Results also display that AMR can provide significantly better performances than GSM FR in the full range of test conditions, and significantly better performances than the GSM HR codec down to 7 dB C/I. The four highest codec modes (7.95, 7.4, 6.7 and 5.9) were found significantly better than the GSM FR in error free conditions down to 13 dB C/I and, at least equivalent to the EFR at 10 dB C/I down to 16 dB C/I. The three highest modes (7.95, 7.4 and 6.7) are equivalent to the error free EFR in very low error conditions. In addition, the two lowest modes were found at least equivalent to the GSM FR over the full range of test conditions.

According to 3GPP TS 26.103 version 4.1.0 Release 4 (section 4, table 4.1), UMTS AMR should be equivalent to GSM AMR FR. Therefore the results of the performance characterization above demonstrated can be accounted considering that information. This is found most relevant for interpretation of this work's results.




| | No Errors | C/I=19 dB | C/I=16 dB | C/I=13 dB | C/I=10 dB | C/I= 7 dB | C/I= 4 dB |
|---|---|---|---|---|---|---|---|
| EFR | 4.21 | | | 4.21 | 3.74 | 3.34 | 1.58 |
| 7.95 | 4.11 | 4.04 | 3.96 | 3.37 | 2.53 | 1.60 | |
| 7.4 | 3.93 | 3.93 | 3.95 | 3.52 | 2.74 | 1.78 | |
| 6.7 | 3.94 | | 3.90 | 3.53 | 3.10 | 2.22 | 1.21 |
| 5.9 | 3.68 | | 3.82 | 3.72 | 3.19 | 2.57 | 1.33 |
| 5.15 | 3.70 | | 3.60 | 3.60 | 3.38 | 2.85 | 1.84 |
| 4.75 | 3.59 | | 3.46 | 3.42 | 3.30 | 3.10 | 2.00 |
| FR | 3.50 | | | 3.50 | 3.14 | 2.74 | 1.50 |
| HR | 3.35 | | | | 3.24 | 2.80 | 1.92 |

**Figure 4.7:** *Family of curves for Experiment 1b (Clean Speech in Half Rate)*

# 5
# Implementation of the Thesis Project

In this chapter, every aspect related to the developed work is presented.
First, an overview of the Project is made, alongside with the summarized objectives. Then, all the requirements that need to be met are presented, followed by the Selection Phase and Laboratory Set-Up, and by the final System Architecture.
Furthermore, a section with all the tests and evaluations is presented, followed by the final results and performances along with the respective interpretations.
The reader will find some mandatory alterations executed during the testing procedure that are due to the various obstacles found while trying to achieve the best results.

## 5.1 Project

The proposed goals for this work consist in characterizing speech radio channels in 2G and 3G. More specifically, assessing the speech quality in clean speech and error conditions (RF noise added to the downlink of the transmission channel), between similar bit rate codecs for different network technologies (2G versus 3G). The sets of similar bit rate codecs for the test cases will be referred as High Rate and Low Rate, for the highest and lowest bit rate codecs, respectively.
Regarding the evaluation of the speech quality, it was defined that the ITU-T P.800 Mean Opinion Score would be the reference metric.
Also, a comparison of the transmission parameters, RxQual and Ec/$N_0$, with the

MOS results obtained, is to be carried out. The first is due to the need to confirm studies by several investigators and, the latter is because no study that correlates the two data was currently found. For a better comprehension, the reader can find in Figure 5.1 an illustration of the proposed tests for the Project:

| | Voice CODECs | | | | | |
|---|---|---|---|---|---|---|
| Noise Levels | GSM | | | UMTS (AMR-NB) | | |
| | FR 13 kbit/s | EFR 12,2 kbit/s | HR 5,6 kbit/s | A 12,2 kbit/s | F 5,9 kbit/s | G 5,15 kbit/s |
| No Noise + Various noise levels | Voice Quality - MOS | | | | | |
| | Full Rate Comparison | | | | Half Rate Comparison | |
| | MOS Vs. RxQual | | | MOS Vs. Ec/No | | |

**Figure 5.1:** *Project Proposed tests*

### 5.1.1 Requirements

To be able to accomplish the objectives proposed for this work, some requirements need to be met. The three main considered requirements are:

- Multiprotocol Radio Communication Tester: one which is able to simulate a mobile cell for the radio networks in test, 2G and 3G. Also, obviously equipped with the necessary licenses for RF analysis of the transmission channels and, with the speech codec options: GSM FR, GSM EFR, GSM HR and UMTS AMR-NB (12.2, 5.9, 5.15 kbit/s);

- Wideband Radio Frequency Generator: needed to generate a pseudo-random RF signal which is to be combined with the signal generated by the emulated cell, thus simulating error conditions in the transmission channel. If possible, with characteristics which allow generation of similar radio frequencies as those found in 2G and 3G cell wireless networks (concerning, for example, digital modulation and bandwidth);

- Speech Quality assessment tool: an intrusive perceptual objective method compliant with MOS metric (e.g. ITU-T PESQ), to compare between original (clean) speech samples and their correspondent degraded samples which go through the cell in test. Preferably employing a Hardware Audio Analyser equipped with the intended software license or, by Software that implements the method's algorithm.

Apart from the above mentioned main requirements for the implementation, other testing materials are also of foremost importance, such as:

- Testing Phone: with known well-functioning radio interface for both 2G and 3G. Preferably with a compatible external antenna connector, and phone headset (with proper connector to link with the speech quality assessment tool);

- Appropriate cabling and connectors: cabling should be properly isolated in order to avoid surrounding EMI radiated and, show low-to-negligible attenuation values. Connectors depend on the materials used;

- Broadband Combiner (passive element): to combine the RF network signal with the RF noise. Should have a frequency range that can be applied to the 900, 1800, 1900 and 2100 MHz bands;

- Broadband Diplexer/Duplex Filter (passive element): allowing the uplink and downlink frequencies to share a common communication channel, and in consequence to block other unwanted frequencies;

- Speech Samples: recommended by the ITU-T for end-to-end speech quality evaluation. Preferably in English language.

Throughout the work some new materials were added to the experiment, as they were needed to continue the work. The reader will find those introduced along the succeeding sections.

### 5.1.2 Selection Phase and Laboratory Set-Up

The selection of the materials is not an easy task since before being put to use they need to be studied in order to prove their feasibility for the work in question. Nevertheless, owing to Óbuda University (Institute of Communications) resources, most of the requirements were met successfully.
The selection phase started with the necessary Multiprotocol Radio Communication Tester:

✓ Rohde & Schwarz CMW500 Wideband Radio Communication Tester (Figure 5.2).

**Figure 5.2:** *Rohde & Schwarz CMW500 Wideband Radio Communication Tester*

R&S CMW500 device is able to test the air interface of wireless devices and, can be used in all phases of product development and production. It supports all common cellular and non-cellular wireless technologies:

- Satellite Navigation: GPS;
- Cellular: LTE FDD/TDD, Mobile WiMAX, TD-SCDMA, CDMA2000 1xRTT, CDMA2000 1xEV-DO, WCDMA/HSPA, HSPA+, GSM, GPRS, EDGE, EDGE Evolution;
- Broadcast: DVB-T, FM stereo, CMMB, T-DMB, MediaFLO;
- Wireless Connectivity: WLAN a/b/g/n, Bluetooth.

This device is capable of performing RF and protocol tests to end-to-end application tests. It is found to be a great solution for the project since it implements all protocol layers, therefore making it possible to alone be used to simulate the intended network cell without any need for further resources.
Regarding the objectives and requirements for the 2G and 3G voice tests, the following device licenses are mandatory. For GSM (Table 5.1):

**Table 5.1:** *R&S CMW software mandatory options for GSM implementation tests [8]*

| Type | Designation |
|---|---|
| R&S CMW-KS200 | GSM GPRS EDGE R6, basic signalling |
| R&S CMW-KS210 | GSM GPRS EDGE R6, advanced signalling |
| R&S CMW-KG200 | GSM GPRS EDGE R6, generator, downlink |
| R&S CMW-KM200 | GSM GPRS EDGE R6, TX measurement, uplink |
| R&S CMW-KM012 | TX measurement, multi evaluation list mode |

**Table 5.2:** *R&S CMW software mandatory options for WCDMA implementation tests [9]*

| Type | Designation |
|---|---|
| R&S CMW-KS400 | WCDMA R99, basic signalling |
| R&S CMW-KS410 | WCDMA R99, advanced signalling |
| R&S CMW-KG400 | WCDMA R99, generator, downlink |
| R&S CMW-KM400 | WCDMA R99, TX measurement, uplink |
| R&S CMW-KM012 | TX measurement, multi evaluation list mode |

For WCDMA, the mandatory device licenses can be consulted in Table 5.2: With these licenses installed on the R&S CMW500 all the requirements are met.

Second on the selection phase comes the Wideband Radio Frequency Generator:

✓ Rohde & Schwarz Signal Generator SME 03 (5 kHz to 3.0 GHz) (Figure 5.3).

**Figure 5.3:** *Rohde & Schwarz Signal Generator SME 03 (5 kHz to 3.0 GHz)*

The R&S SME03 supplies the complex signals required for the development and testing of digital radio receivers. It is capable of generating signals used in some of the main digital radio networks in line with relevant standards regarding the type of modulation, data format, TDMA structure and frequency hop patterns.
Although R&S SME03 is not the most up to date device, it is still able to generate the RF signals needed to meet the demands. It presents the following relevant capabilities:

- Radio Frequency range from 5 kHz to 3.0 GHz, this way comprising the 900, 1800, 1900, 2100 MHz bands;
- Digital Modulation: Pseudo Random Binary Sequence (PRBS) source generator, GMSK and Quadrature Phase Shift Keying (QPSK) modulations needed to emulate GSM and WCDMA signals;

- Radio Frequency Sweep: with START, STOP, CENTER and SPAN options, this way allowing a bandwidth on the RF generated (once again to emulate a typical cell network signal).

RF connections on both R&S devices are 7/16 female N type connectors, with 50 Ω nominal input impedance.

Regarding the Speech Quality assessment tool, first an attempt was made to arrange a Hardware/Software combination, which is the professional approach (e.g. from network operators) used for intrusive perceptual objective speech quality assessment. Since the University did not own one, contacts were made to various companies concerning this matter. Although having found an available R&S UPV Audio Analyser (for a trial usage period), licenses for ITU-T PESQ (UPV-K61) and POLQA (UPV-K63) were not included and so, it was not possible to use the device to measure the speech quality.
With a limited time to work on the project, waiting for a device with the licenses needed either from Rohde & Schwarz, Cisco Systems or other similar companies was out of the question. Consequently a new approach had to be made so that the assessment of speech quality could be performed.
After some research on ITU's webpage, more specifically on P.862 (Amendment 2) and P.863 Recommendations, it was found that P.862 annex A (2005) had a compilable source code (therefore executable) of the PESQ software. POLQA's Recommendation, on the other hand, only provided for parts of the source code representing the algorithm utilized for it, and not a compilable (or executable) version. This way ITU-T PESQ tool, and according to its Intellectual Property Rights, is permitted for usage in this project's tests.
Moreover, whilst searching and contacting for the above-mentioned Hardware/Software combination on various companies, a computer software tool that could perform MOS and PESQ MOS measurements was found, Sevana Oy's AQuA. This is worth mentioning since Sevana Oy was the only company which answered positively to the request of using their software for this experiment.
At this point it was defined that both ITU-T PESQ and Sevana Oy AQuA software tools would be the ones applied to assess speech quality.

Having the tools for the evaluation, new questions surfaced. As the software tools are not part of a hardware audio analyser, they have to be executed in a computer. Thus, a recording method for the speech samples that will be under test is necessary.

Since these tools are intrusive objective methods of measuring quality, original (clean) samples have to be played into the phone, go through the emulated network and then, have the degraded echoed speech samples recorded back. This way, enabling a comparison between the original and degraded samples (executed by the software tools). This system is illustrated in Figure 5.4.

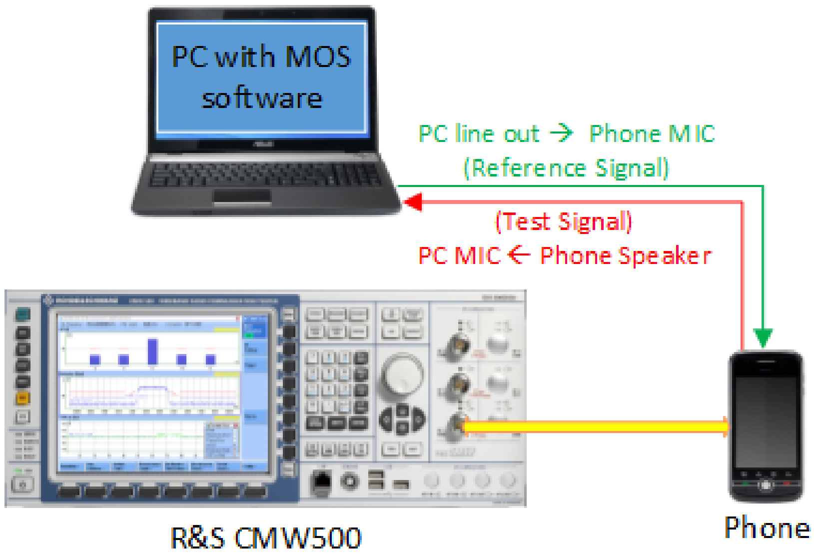



**Figure 5.4:** *Illustration of the basic system configuration*

As shown in Figure 5.4, the phone headset connection is to be used to capture the Reference Signal coming from the Personal Computer's (PC) line out into the Phone's Microphone (MIC) and then, looped back from the Phone Speaker line into the PC's MIC. Although seeming an easy plug and play configuration, impedance and signal level differences between computer/phone constitute a problem. Consequently, impedance adaptation and signal level matching is required.
Before solving this impedance problem, the Phone that was going to be under test had to be selected. Having no access to a phone with an external antenna connector, the choice pended for a personal white label smartphone which had been under successful usage for the last two and half years. This phone never had any problems whatsoever and, has always shown a great network radio reception both in GSM and WCDMA. The general specifications of the Gigabyte GSmart G1305 Boston smartphone can be found on Table 5.3.

To solve the problematic of the impedance and signal level difference one could go either with a passive or active adaptation element. After analysing the benefits and drawbacks of both approaches, the choice clearly inclined for the active element. Mainly because an active element would automatically act upon variations

**Table 5.3:** *Phone Under Test general specifications [10]*

| | | |
|---|---|---|
| | 2G Network | GSM 850 / 900 / 1800 / 1900 |
| **Networks** | 3G Network | HSDPA 900 / 1900 / 2100 |
| | GPRS | Class 10 (32 - 48 kbit/s) |
| **Data** | EDGE | Class 10 |
| | HSDPA speed | 7.2 Mbit/s |
| **Audio** | Headset | 3.5 mm jack connector |
| | Operative System | Android OS, v2.2.2 |
| **Features** | Chipset | Qualcomm MSM7227 |
| | CPU | 600 MHz ARM 11 |

in the signal level, not being stuck to an impedance and voltage level adaptation. This shows some big advantages when compared with a passive element such as the greater minimization of noise, distortion and ground loops.
Once again the University was found to be resourceful since an Active Direct Injection (DI) unit was available for use. The DI unit provided is: Behringer ULTRA-DI Model DI100 (Figure 5.5) and, the relevant technical specifications can be seen on Table 5.4.


**Figure 5.5:** *Behringer ULTRA-DI Model DI100 (Active Direct Inject Box)*

DI units are usually required for stage and studio, or other music instrument related applications, whether professional or not. These units main purpose is to provide impedance and signal matching, for example, from a high impedance and unbalanced signal and then converting it to a low impedance, balanced signal. Which is the case seen in the illustration of the basic system configuration (Figure 5.4). The reference signal coming from the high impedance and unbalanced signal on the PC's line out, which needs adapting to be able to be used by the Phone's Microphone.

**Table 5.4:** *Behringer ULTRA-DI Model DI100 specifications [11]*

| | |
|---|---|
| **Frequency Response** | 10 Hz to 93 kHz |
| **Noise Level** | -110 dBu |
| **Total Harmonic Distortion + Noise (@ 1 kHz / 0 dBu)** | EDGE <0.005 % |
| **Input Impedance** | >250 kΩ |
| **Load Impedance** | >600 Ω |
| **Inputs** | 6.35 mm jack and XLR (unbalanced) |
| **Maximum Input level** | +10/ +30/ +50 dBu |

At this point, almost all the proposed requirements were met. Still, a Broadband Combiner and Diplexer, appropriate cabling/connectors and, Test Speech Samples in conformance with ITU-T speech quality evaluation standards, were necessary.
A Broadband Diplexer was not available, though the needed Broadband Combiner was provided by Vodafone Hungary. Vodafone was not only able to arrange for the combiner for use in this project, but also two professionally isolated cables with negligible loss. These cables being exactly suited for wireless cell network tests, with 7/16 male N type connectors at the cables endings. The broadband (2 way) direction coupler presents low dissipation losses and VSWR. The relevant characteristics for the BHLT10 tri-band Power Splitter/Combiner are present in Table 5.5.

**Table 5.5:** *BHLT10 tri-band Power Splitter/Combiner - Characteristics*

| **Frequency Band** | 880 to 960 MHz | 1710 to 2170 MHz |
|---|---|---|
| **Reflection** | <-20 dB | <-20 dB |
| **Isolation** | >25 dB | >25 dB |
| **Nominal Coupling Loss** | 3 dB ± 0.2 dB | 3 dB ± 0.3 dB |
| **Connectors** | N - female | |

As it can be checked in the characteristics and, also by taking into account the fact that this is a combiner utilized in wideband radio tests on wireless cell networks, all the characteristics meet the requirements.

Given that the available phone for the experiment did not have the means for an external antenna cable connection, the air interface had to be used to connect to the R&S CMW500 emulated cell. So, an indoor Monopole Antenna with the following characteristics was used:

- Directivity: Omni-directional;
- Polarity: Vertical;
- Application/Usage: for all VHF and UHF frequencies;
- Gain: 0 dBi.

Concerning the cables and connectors, only the Phone/PC, and DI unit were in need of connection links. As these needed cables are not easy to find, they had to be handmade produced. The required pieces to create these cable connection are:

- Two 3.5 mm jacks (Mono): PC MIC input and line out;
- One 3.5 mm jack (Stereo + Mic): Phone headset connection;
- One 6.35 mm jack (Mono): unbalanced input on the DI-unit;
- One XLR female connector: balanced output on the DI-unit;
- Connection cables.

After a study of the correct pin-out for the connectors and the respective device connections (with the correspondent datasheets), plus with help of a multimeter, a schematic for the soldering process was designed and can be consulted in Figure 5.6.

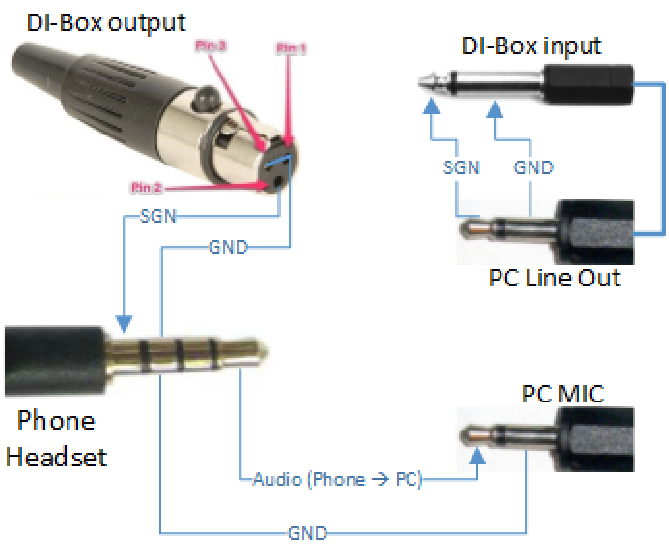



**Figure 5.6:** *Schematic of connections for Phone, PC and DI-Box*

With this configuration the last cabling requirements were set.

Lastly, the needed Speech database for the experiment was downloaded from the ITU-T P Series Test Signals for Telecommunication Systems webpage, more specifically from ITU-T P.50 Recommendation. The set of "real speech recordings" found in IUT-T P.50 Appendix I were recorded in the laboratories of some ITU members and include, 16 sentences in each 20 languages. The samples are about 11.2 s in length and active on 83 % of the time, on the average. Note that for each set (or subset), half is male and the other half in female talker records. These test samples were especially created for telephonometry applications and it is in conformance with the standard demands for speech evaluation. The characteristics of the speech samples are: Wave format (16-bit linear PCM, 44-byte header), Mono, at 16 kHz sample rate.

Adding to all requirements, one which is not stated as needed but it can be considered quite useful in network testing, is a proper Subscriber Identity Module (SIM) test card. The R&S CMW-Z04 - Mini Universal Integrated Circuit Test Card was provided, therefore relegating the need to use a random SIM card (operator locked), as this Mini UICC Test Card ensures total compatibility and reliability to work with the R&S CMW500.

### 5.1.3 System Architecture

The System Architecture illustrated in Figure 5.7, shows the final structure of the *montage* with all the main components that are a part of the experiment. As it is possible to verify, the arrows in the interconnections represent the direction of the interactions between the various components of the system.

The system architecture consists in using the Wideband Radio Communication Tester R&S CMW500 to emulate the wireless network cell for each test. Also, allowing the measurement of transmission parameters (e.g. the ones in test, RxQual and $Ec/N_0$) that are very useful in monitoring and interpreting the behaviour of the call in progress.

The Signal Generator R&S SME03 is in turn used to inject different RF noise levels into the downlink of the communication channel. This is done with the purpose of degrading the conditions in which the calls are made, for subsequent analysis.

The Broadband Combiner is the responsible element for combining both network and noise signals injected in it, and then transmitting the resulting signal via air interface, with the aid of an Omni-directional indoor antenna.

The PC in Figure 5.7, is in charge of two main roles in the system architecture. In

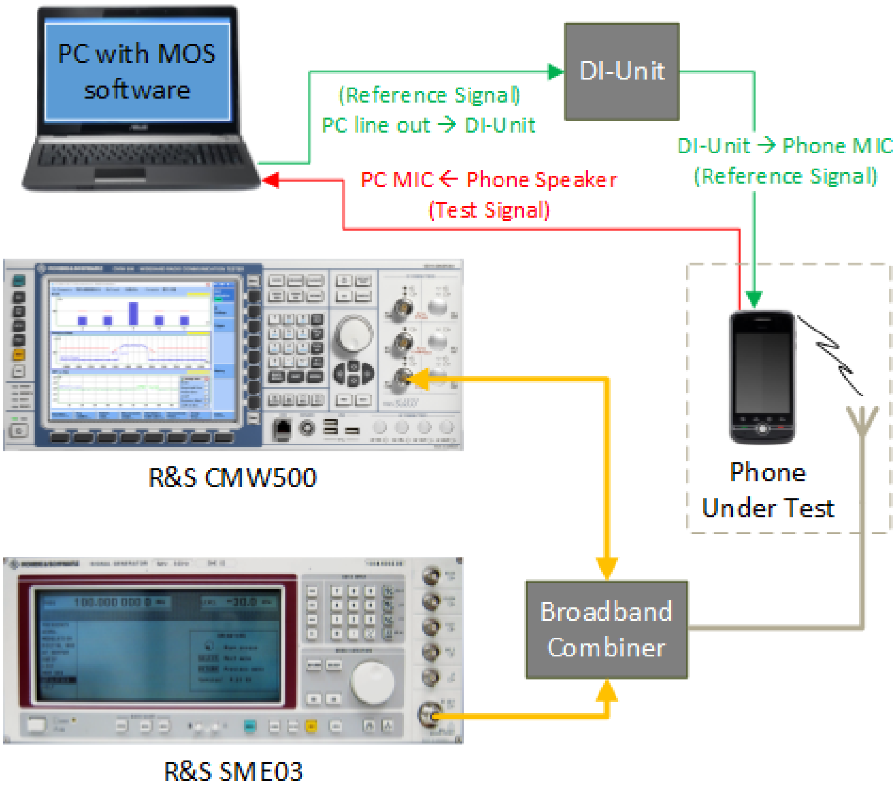



**Figure 5.7:** *System Architecture*

a first step, to Play and Record the test speech samples and, in a second and final stage, to assess the speech quality employing the intrusive objective method chosen (ITU-T PESQ MOS, and Sevana Oy AQuA PESQ MOS). The reference signal is sent through the PC line out into the DI-Unit. The signal coming from the PC line out is unbalanced and so it passes through the DI-Unit with the purpose of making the impedance and signal matching needed for the Phone's MIC input. On the other hand, the degraded test signal that is echoed from the network, back to the Phone Under Test, into the PC MIC, does not need this impedance and signal adaptation since both PC Speaker and PC MIC are of the same balanced levels.

## 5.2 Testing and Evaluation

In the following sections the reader will find a description of all the tests performed, corrections and validations corresponding to each part of the work. Along with the evaluation and interpretation of the results.

The main sections of this chapter encompass all the procedures from the Configuration and Calibration of devices and parameters, to the Speech Quality Evaluation and subsequent comparison with relevant transmission parameters. Finalizing with a section for the discussion of results.

### 5.2.1 R&S CMW500 Configurations

Before any test execution, one has to make sure the configurations used are not only optimized for the best results, but also that the conditions are the same (or the closest to the same) for each and every test. This is done with the obvious intent of making the final results as reliable as possible.
The configurations for the R&S CMW500 Wideband Radio Communication Tester should check the following items:

- ✓ Channels/Bands: since the air interface is utilized, the frequencies have to be free for use. This meaning that those frequencies cannot be allocated by any network operator or entity of the same sort;

- ✓ Cell Transmission Power (output power): approximately the same for both 2G and 3G tests;

- ✓ Data Source: this has to be set for echo/loop mode, in order to get the network echoing the test samples back to the phone;

- ✓ Miscellaneous parameters: attenuation/gain on RF input/output of the device, set to none. Switching off the Packet-Switched domain, as only the Circuit-Switched is needed for voice, among others.

Regarding the unallocated (not in use) frequencies for the experiments, the information was provided by Vodafone, with the insurance that the following Channels/Bands were free for use 5.6:

**Table 5.6:** *Unallocated (free for use) frequencies*

| | | | |
|---|---|---|---|
| **GSM** | Channel/Band<br>121 and 122 | | |
| **UMTS (FDD)** | Band 1 | UL freq.: 1965 - 1980 MHz | DL freq.: 2155 - 2170 MHz |

For GSM the channel 122 (DL 959.4 MHz) was used for the BCCH, and 121 for the TCH/PDCH (DL frequency of 959.2 MHz and UL frequency of 914.2 MHz). As for UMTS, the channels follow the rule found in 3GPP TS 25.101, section 5.3.a, which state that in Band 1 the Tx-Rx frequency separation for UTRA FDD has to be of 190 MHz. Thus, the selections for UMTS were:

- Downlink: channel 10823 (2164.6 MHz);
- Uplink: channel 9873 (1974.6 MHz).

It was defined the Power Control Level (PCL) would be 12 in value, which means a Nominal Output power of 19 dBm (for GSM 900) [3GPP TS 45.005, table 4.1-4a], and the Downlink Reference Level of -20 dBm. The same output power configurations were made for UMTS. The Tx power levels were then confirmed during active calls with the multi evaluation option of R&S CMW500, for both GSM and UMTS. The data source was configured appropriately, echoing the tests samples. The 'external attenuation' option of the RF input/output was set for 0 dB. This way, not interfering with the signal power. Also, the packet-switched domain was switched off since there is no need for it.

### 5.2.2 Calibration of Volume Parameters

As said before, the employed method for the recording of the tests samples in the experiment is not the standard way of performing it (not the professional way done in companies and organizations where a Hardware/Software combination, e.g., Audio Analyser device plus a PESQ license, is used). For that reason, and since the volume and gain parameters utilized in the PC line out, PC MIC and Phone, can have a big influence in both the reference and respective degraded samples, a study phase to discover the optimized values for these parameters was mandatory.

The parameters that can be tuned are the PC line out volume, PC MIC volume, PC MIC gain and the Phone speaker volume.
Right away the PC MIC gain can be set to the correct value for the recordings, 0 dB. The reason for no need in gain is based on the principle that the gain would not only increase the signal strength, altering the signal, but also the noise associated in the transmission.
Then, and before any test, the configuration of the Play and Record parameters for

the samples was executed according to the ITU-T PESQ specifications:

- Play: Original samples, Mono;
- Record: Wave format (16-bit linear PCM, 44-byte header), Mono.

The sample rate of the Play and Record was set at 44 kHz due to inability of the hardware to record at lower sample rates, having later, both original and degraded samples converted back to 8 kHz for the speech quality evaluation.
In order to optimize volume levels, a set of recordings were made for both GSM and UMTS technologies. The objective was to approximate the spectrum of the degraded signal to the spectrum of the correspondent original signal. The tests were obviously carried out in clean speech conditions (no noise added). After an extensive number of tests, the best results for both GSM and UMTS were found to be:

✓ PC Line Out volume: 35 (out of 100);

✓ PC MIC volume: 38 (out of 100);

✓ Phone Speaker volume: 2 (out of 5).

Figure 5.8 illustrates an example of a recording performed with these configurations, compared with its original signal, in GSM.

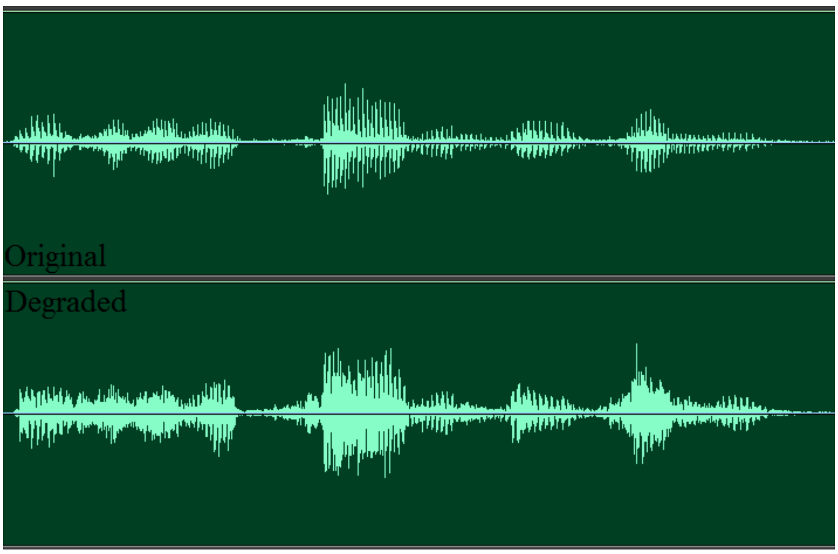



**Figure 5.8:** *Recording example (for GSM) with volume parameters optimized - Comparison between reference original signal (above) and respective degraded signal (below)*

It is possible to verify clear differences in amplitude which in turn increase the Root Mean Square (RMS) SNR value. Also some distortion is visible on the spectrum. These aspects will reflect negatively in the speech quality evaluation. The distortions can be mostly justified by the fact that the PC audio card is not suited for audio analysis such as these. Also, the losses in the cabling between the Phone and the computer should be accounted in that matter.

Although in practice, it is not very relevant to the speech quality assessment algorithm, since this is corrected during its execution, a difference in the initial echo delay (not visible on the recording example) was noticed between the GSM and the UMTS recordings. An echo delay is expected, since the samples have to go through the simulated cell on R&S CMW500, back to the PC. Therefore having all the associated delays of the speech transmission process portrayed in Chapter 3.1 and of the transmission through the cables.
This initial echo delay difference between 2G and 3G is quite relevant and, after 36 tests for each technology the weighted average initial delays were:

- GSM: 1.3487 s
- UMTS: 0.8183 s

Having done the tests with the proposed codecs in evaluation (GSM FR, EFR, HR, and UMTS AMR-NB 12.2, 5.9. 5.15), it is believed that this initial delay is caused by differences in the way the 2G and 3G technologies process and transmit data through the communication channel. Not believing in any codec relation for the initial echo delay, as for example, GSM EFR algorithm (ACELP) is the same as the one used in UMTS AMR-NB and yet, no difference is noted in delay between EFR and the other GSM codecs.

On another note, and after a close naked eye inspection to the spectrum of recorded degraded samples a variable delay during the speech sentences was also noticed, but only for the 3G. This can be visualized in Figure 5.9.

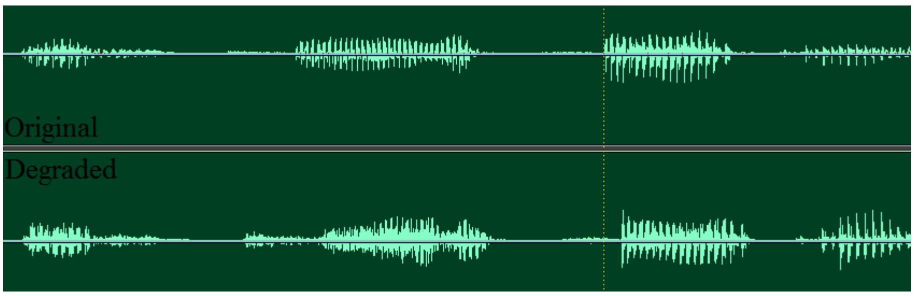



**Figure 5.9:** *Variable delay in UMTS*

This may be explained by the same reasons stated for the initial echo delay, since different technologies utilize different ways to transmit and receive information.

### 5.2.3 Tests - Clean Speech

Having all the devices configured and ready for the experiment, the first step was to begin with the recordings in clean speech (no noise) and test both chosen speech quality measuring methods, ITU-T PESQ and AQuA.

From the speech database in hand, the chosen language for testing was American English in 4 different samples (2 male + 2 female) all different speakers (subjects). Also, for each test case (codecs and respective samples) a significant number of recordings was performed, with the intent of having the most credible final results. Three recordings of each 4 samples, making a total of 12 samples for each codec, and a total of 36 for each technology, in each test condition.

Once a full set of recordings was completed in clean speech conditions, these were analysed in ITU-T PESQ and AQuA PESQ. With the first results gone through the algorithms, it was noticed that the MOS values were below the real values expected (see Chapter 4.3 - 3GPP's Performance characterization (clean speech results)). The approximate expected MOS values can be consulted on Table 5.7:

**Table 5.7:** *Expected average MOS values in Clean Speech (3GPP's Technical Report 26.975 section 5)*

| GSM FR | GSM EFR | GSM HR | AMR-NB 12.2 kbit/s | AMR-NB 5.9 kbit/s | AMR-NB 5.15 kbit/s |
|---|---|---|---|---|---|
| 3.5 | 4.01 | 3.35 | 4.01 | 3.72 | 3.5 |

The MOS results from the tests were below the expected for both ITU-T PESQ and AQuA PESQ. The ITU-T PESQ results presented even lower values when compared with AQUA PESQ and, did not show usable results for the WCDMA recorded samples. For example, the samples recorded for WCDMA AMR 12.2 kbit/s had lower MOS value than GSM EFR and FR (which is not realistic). On the other hand the AQuA PESQ algorithm showed a proportional difference closer to what was expected for both 2G and 3G. For this reason the ITU-T PESQ algorithm was considered to be put aside from the experiment, having only AQuA PESQ being used as a solution for the speech quality assessment from this point forward.
The reasons found to explain these strange results with ITU-T PESQ are:

- As stated in 3GPP TS P.862, in relation to the downloadable ITU-T PESQ software, conformance tests should be performed with a specific set of samples, the ITU-T P-series Supplement 23;

- Recognized inaccuracies with wideband scenarios and CDMA codecs;
- Unreliable and not recommended for use with long speech samples, e.g. 12 s, which is the case for the speech database chosen (the conformance samples annexed with P.862 are of approximately 6 s).

Measured AQuA PESQ MOS values for clean speech can be consulted in Figure 5.10:

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| F1 | 3,20 | 2,95 | 3,05 | 3,35 | 3,08 | 2,96 |
| | 3,23 | 2,85 | 2,80 | 3,42 | 3,04 | 3,16 |
| | 3,16 | 3,18 | 3,18 | 3,38 | 3,01 | 3,04 |
| F3 | 3,01 | 3,16 | 3,37 | 3,11 | 3,03 | 2,94 |
| | 3,01 | 3,12 | 2,79 | 3,29 | 3,03 | 3,08 |
| | 3,01 | 3,39 | 3,19 | 3,12 | 3,10 | 3,05 |
| M1 | 2,79 | 2,83 | 2,83 | 3,17 | 2,63 | 2,65 |
| | 2,80 | 2,75 | 2,76 | 3,02 | 2,62 | 2,69 |
| | 2,86 | 3,06 | 2,64 | 3,17 | 2,62 | 2,70 |
| M3 | 2,90 | 3,15 | 2,61 | 3,16 | 2,74 | 2,71 |
| | 2,87 | 2,69 | 2,32 | 2,96 | 2,66 | 2,85 |
| | 2,85 | 2,97 | 2,62 | 2,86 | 2,76 | 2,86 |
| **Average Total** | **2,97** | **3,01** | **2,85** | **3,17** | **2,86** | **2,89** |
| **Standard** | **0,15** | **0,21** | **0,30** | **0,17** | **0,20** | **0,17** |
| **Average F1, F3** | 3,10 | 3,11 | 3,06 | 3,28 | 3,05 | 3,04 |
| **Average M1, M3** | 2,85 | 2,91 | 2,63 | 3,06 | 2,67 | 2,74 |

**Figure 5.10:** *AQuA PESQ MOS values (Clean Speech)*

For convenience, UMTS AMR-NB codecs will be given class letters, as it is done on the R&S CMW500 device. So, AMR-NB A, F and G, are of 12.2, 5.9, 5.15 kbit/s, respectively. Additionally, the names for each sample were given in accordance to the original order in the downloaded speech database, so F1 and F3 are the first and third female record samples, M1 and M3 are for the respective males.

First of all, the big MOS difference, [0.43 ; 1.00] less when compared with the real values, can be easily explained due to the recording method employed, with which the chances for degradation are high. As mentioned before, the "professional" way of perceptual objective evaluation of speech quality is executed with an adequate hardware audio analyser which implements the speech quality algorithm. The degradations in the hand-made cables, the not perfect optimization of volume configurations (Chapter 5.2.2), the audio sound card on the PC not being optimized for this kind of applications, these are all sources of speech quality degrading.

Notwithstanding, the behaviour between codecs is similar to the expected.
In clean speech conditions the AMR-NB 12.2 kbit/s (A) shows the best quality followed by the other two highest bit rate codecs, the GSM FR and EFR, as it was predictable. Since in these conditions the probability for errors is minor and more information reaches the destination untouched.
GSM HR codec demonstrates the highest standard deviation, 0.3, therefore it is the most unreliable result, though it still confirms it as the lowest score, followed by the other two low bit rate codecs. With no statistically relevant difference among them.

Aside from the MOS values, transmission parameters for the testing were also recorded. The RxQual for 2G tests was always 0 (the best result possible), indicating the low bit error rate found in the clean speech connection. For the CPICH $Ec/N_0$ the values were -2.5 dB / -2 dB (lower/upper values), also very close values to the best possible, 0 dB.

AQuA algorithm also delivers results for the Voice Quality in percentile values. This value represents the signal similarities in terms of quality between the degraded and the original samples. The clean speech voice quality percentile values can be consulted in Figure 5.11.

| | **FR** | **EFR** | **HR** | **AMR-NB A** | **AMR-NB F** | **AMR-NB G** |
|---|---|---|---|---|---|---|
| **F1** | 83,83 | 77,91 | 80,41 | 87,14 | 80,91 | 78,27 |
| | 84,54 | 75,54 | 74,38 | 88,75 | 80,05 | 82,75 |
| | 82,78 | 83,28 | 83,40 | 87,70 | 79,44 | 80,08 |
| **F3** | 79,50 | 92,87 | 87,56 | 76,66 | 79,81 | 77,77 |
| | 79,52 | 90,86 | 75,13 | 80,82 | 79,96 | 81,11 |
| | 79,47 | 88,03 | 82,25 | 76,83 | 81,38 | 80,39 |
| **M1** | 73,92 | 74,90 | 74,91 | 83,13 | 69,77 | 70,28 |
| | 74,18 | 72,97 | 73,25 | 79,69 | 69,54 | 71,46 |
| | 75,69 | 80,63 | 70,07 | 83,18 | 69,56 | 71,68 |
| **M3** | 76,85 | 82,60 | 69,43 | 92,86 | 72,63 | 71,96 |
| | 76,12 | 71,32 | 61,53 | 78,13 | 70,72 | 75,45 |
| | 75,53 | 78,51 | 69,54 | 75,77 | 72,66 | 75,86 |
| **Average Total** | **78,49** | **80,79** | **75,16** | **82,56** | **75,54** | **76,42** |
| **Standard** | **3,70** | **6,98** | **7,26** | **5,54** | **5,06** | **4,29** |
| **Average F1, F3** | 81,61 | 84,75 | 80,52 | 82,98 | 80,26 | 80,06 |
| **Average M1, M3** | 75,38 | 76,82 | 69,79 | 82,13 | 70,81 | 72,78 |

**Figure 5.11:** *AQuA Voice Quality in Percentage (Clean Speech)*

With these percentage values the MOS results are reinforced. Confirming the codec's behaviour aforementioned.

### 5.2.4 Tests - Error Conditions

For the second and most important part of the experiment, the transmission channel is subject to noise injection in the downlink frequency.
First, a couple of analysis have to be performed with the signal generator (R&S SME03), which will be injecting the noise signal into the system. The set of analysis consist in:

1. Configuration of noise signal parameters: Bandwidth (sweep option), Digital Modulation and Downlink frequency;

2. Determination of RF noise levels for the tests, in all codecs.

First the Bandwidth values were set according to the GSM and UMTS specifications, 200 kHz and 5 MHz carrier bandwidths, respectively. The span of the band with its centre frequency being the downlink used frequency. The digital modulation applied in GSM and UMTS are GMSK and QPSK, respectively. These are common digital modulations used in these technologies, and the device also did not offer any other plausible digital modulations for use.
The GMSK digital modulation was configured as following:

- Source: PRBS 15 bit;

- Bit Rate: 270.833 kbit/s, according to 3GPP Technical Specification 45.004.

The QPSK was set with the subsequent configurations:

- Source: PRBS 15 bit;

- Bit rate: 48.6 kbit/s.

The frequencies used are 959.2 MHz for GSM and, 2164.6 MHz for WCDMA.

For the RF noise power level determination the ITU-T P.800 Listening Quality experiment instructions were used. These relate the listening quality in MOS values with the effort required to understand the meanings of sentences (Table 5.8). Successive recordings were performed, trying to look for the different MOS levels when listening to the degraded samples and, consequently taking note of the power level for each. The increase in noise was performed gradually with a resolution of 0.5 dBm in each step. Obviously, as these depended on the perception of the quality,

**Table 5.8:** *ITU-T P.800 Recommendation - Listening Quality Instructions for MOS values*

| MOS | Effort Required |
|---|---|
| 5 | Complete relaxation possible; no effort required. |
| 4 | Attention necessary; no appreciable effort required. |
| 3 | Moderate effort required. |
| 2 | Considerable effort required. |
| 1 | No meaning understood with any feasible effort. |

the values are subject to error in the process of listening perception.
It was noticed that the quality did not change (did not degrade) when using the sweep option on the signal generator. This can be explained by the way the cell behaves when it detects this noise spread over the width of the signal. The hypothesis is that it detects it as it was a variant of white noise, since it behaves randomly and with a constant level over the whole bandwidth. Therefore being filtered during the reception process in the base station.
The sweep option was turned off, therefore having noise inserted only in the centre carrier frequency and, the tests were repeated. This time the degradation was noticeable and the needed noise levels for the experiment were determined. The RF noise levels to be tested are: -4.5, -3.0, -1.5, 0.5, 2.0 and 3.0 dBm. The increasing in noise level is not proportional since the 'listening effort' required was considered the main reference for the level determination.

The detailed tabled MOS values for each recording and each RF noise level, can be found in Annex A. As for the Voice Quality Percentile values those are presented in Annex B.
The analysis and interpretation of results for, MOS, Voice Quality Percentage, RxQual and $Ec/N_0$ is made in the following section (5.2.5), the Results and Performance sub-chapter.

### 5.2.5 Results and Performance

In this chapter the interpretation of results in clean speech and error conditions is made.
First, the behaviour of speech quality with the increase of noise levels in the transmission channel is analysed. Then, a comparison is made on the performance of codecs with similar bit rates between different wireless cellular technologies (GSM and WCDMA). In the final sub sections a relation between MOS and transmission parameters RxQual and $Ec/N_0$ is explored.

### MOS versus RF Noise

The final average values for the AQuA PESQ MOS, calculated from the values present in Annex A, are shown on Figure 5.12.

| Pn (dBm) | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| no noise | 2,97 | 3,01 | 2,85 | 3,17 | 2,86 | 2,89 |
| -4,5 | 2,96 | 3,00 | 2,86 | 3,23 | 2,85 | 2,88 |
| -3,0 | 2,91 | 3,02 | 2,84 | 3,25 | 2,84 | 2,84 |
| -1,5 | 2,87 | 3,02 | 2,81 | 3,14 | 2,84 | 2,83 |
| 0,5 | 2,75 | 2,97 | 2,70 | 3,12 | 2,84 | 2,82 |
| 2,0 | 2,43 | 2,71 | 2,24 | 2,88 | 2,61 | 2,60 |
| 3,0 | 1,90 | 2,31 | 1,73 | 2,37 | 2,22 | 2,18 |

**Figure 5.12:** *AQuA PESQ MOS Average values (Clean Speech + Noise Levels)*

For better view and understanding of the results of Figure 5.12, these are graphically represented in Figure 5.13.

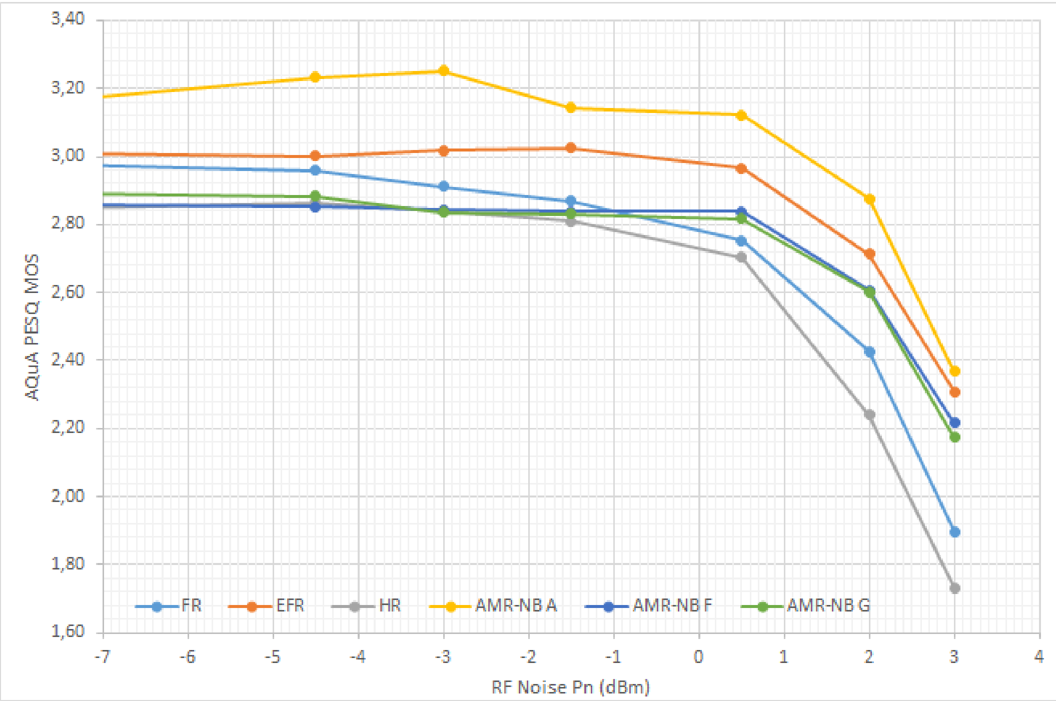



**Figure 5.13:** *Graphic representation of AQuA PESQ MOS Average values (Clean Speech + Noise Levels)*

Note: 'No noise level' in the graphics is represented in the origin of the X-axis (-7 dBm).

By looking at the curves it is possible to verify the speech quality stays statistically unaffected until Pn = -2 dBm. The results demonstrate that GSM FR and HR codecs show less robustness than the other codecs since they show a slight decrease starting at Pn = -3 dBm, when others only start showing quality loss with higher noise levels. FR and HR codecs also show a faster decrease in quality for higher noise levels.
Inversely, codecs with ACELP based algorithms, EFR and UMTS AMR-NB A, F and G show the best resistance to radio frequency noise increase. With AMR-NB A (12.2 kbit/s) having the best quality across the entire tests, followed by EFR.
Although AMR-NB F and G do not start with MOS values as high as other codecs, these two show less lowering in MOS values whether counting from the lowest or the highest noise levels injected. Which is something to be expected from AMR-NB low bit rate codecs, as they are more adequate for noisy transmission situations such as low SNR (being the case).
Taking into account the Performance Characterization study of the AMR codecs, executed by 3GPP members, the increase in quality from a no noise situation for the first noise level injected (- 4.5 dBm), in the AMR-NB A, is expected.

Considering the AQuA Voice Quality Percentage final average values calculated from tables present in Annex B, those can be consulted in Figure 5.14.

| Pn (dBm) | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| No Noise | 78,49 | 80,79 | 75,16 | 82,56 | 75,54 | 76,42 |
| -4,5 | 78,04 | 79,02 | 75,80 | 85,07 | 75,44 | 76,22 |
| -3,0 | 76,91 | 79,42 | 74,37 | 85,70 | 75,24 | 75,08 |
| -1,5 | 75,81 | 79,46 | 74,36 | 82,41 | 75,14 | 74,93 |
| 0,5 | 72,79 | 78,16 | 71,63 | 82,74 | 75,10 | 74,64 |
| 2,0 | 64,72 | 71,76 | 59,69 | 75,82 | 69,33 | 69,69 |
| 3,0 | 50,73 | 62,78 | 46,30 | 63,37 | 61,51 | 58,21 |

**Figure 5.14:** *AQuA Voice Quality in Percentage, Average values, (Clean Speech + Noise Levels)*

If one examines the graphical representation (Figure 5.15) of the values presented in Figure 5.14, the same interpretations as the aforementioned for AQuA PESQ MOS values can be taken. Confirming that are other ways of objective perceptual intrusive speech quality evaluations, not only PESQ MOS (and the most recent POLQA), that can be used with great reliability.

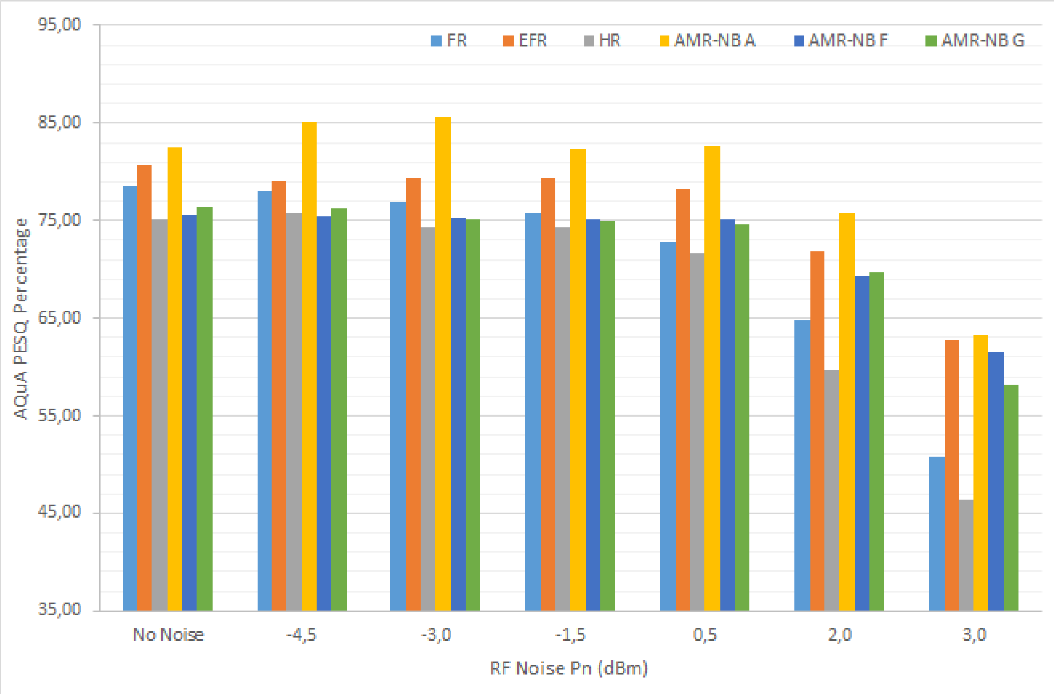



**Figure 5.15:** *Graphic representation of AQuA Voice Quality in Percentage, Average values, (Clean Speech + Noise Levels)*

### Comparison on Codecs with similar bit rates

One of the main objectives of this project is to compare the performance of codecs with similar bit rates between different technologies (GSM and WCDMA). Which are termed as High Rate and Low Rate, for the highest and lowest bit rate codecs, respectively.
In order to go into more detail concerning the results shown in Figure 5.13 - Graphic representation of AQuA PESQ MOS Average values (Clean Speech + Noise Levels), the High Rate and Low Rate codecs were separated subsequently: Figure 5.16 and Figure 5.17.

For the High Rate codecs it is noticed the difference in quality between the UMTS AMR-NB A codec and the other two GSM codecs. Though the GSM ACELP based algorithm, EFR, shows a greater resilience to noise increase when compared with FR. From this it is possible to assert that in similar bit rates, the WCDMA codec is better in quality and it is indeed more robust. Having in mind that AMR-NB tests were executed with fixed rates, the main difference noticed outcomes from little details in the conception of the transcoding functions of the AMR-NB codecs.

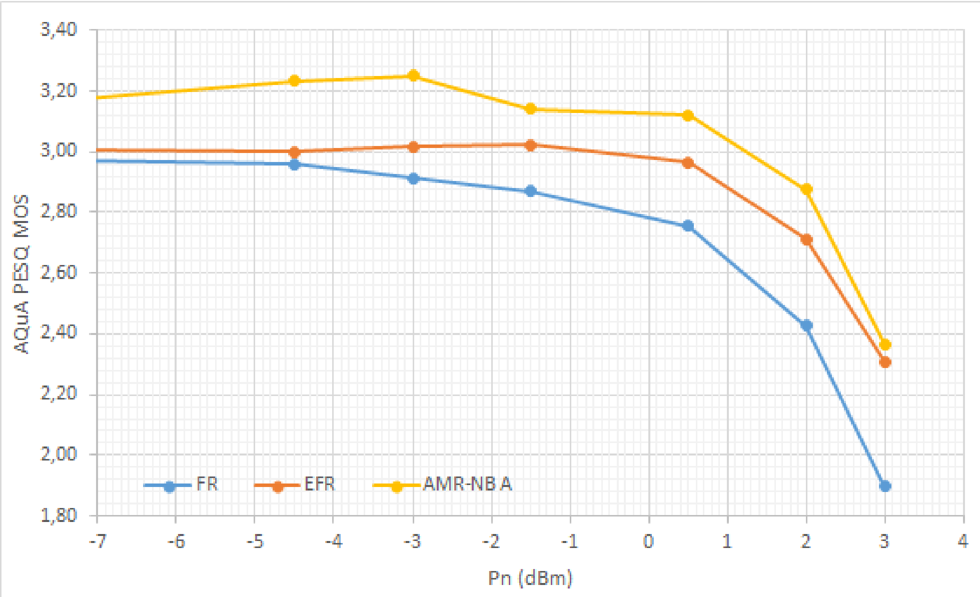



**Figure 5.16:** *AQuA PESQ MOS Average values (Clean Speech + Noise Levels) - High Rate codecs*

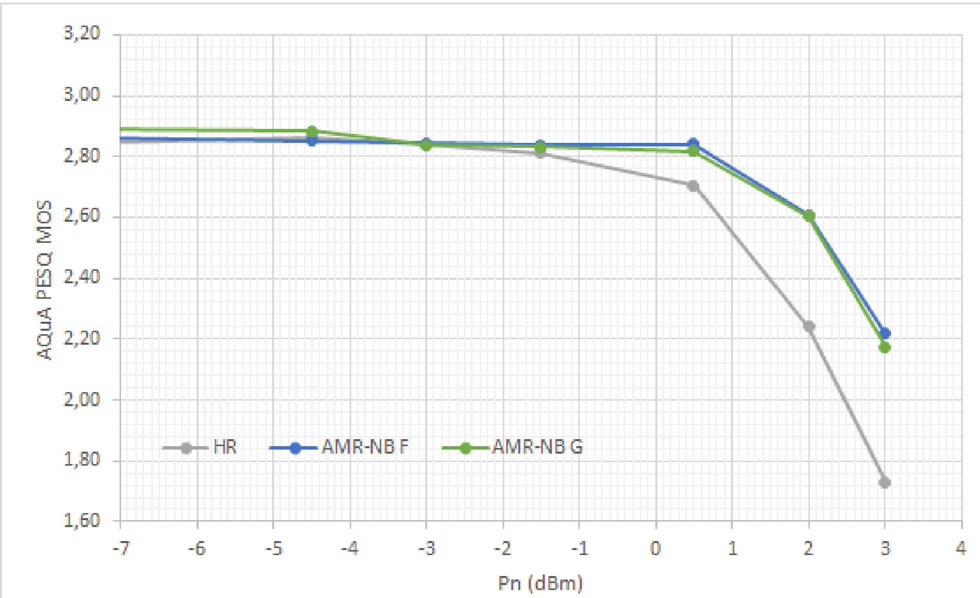



**Figure 5.17:** *AQuA PESQ MOS Average values (Clean Speech + Noise Levels) - Low Rate codecs*

For the Low Rate codecs the difference is not noticed in clean speech conditions but, as SNR decreases, starting on -1.5 dBm, the GSM HR falls short in comparison with the AMR-NB codecs. Once again, the WCDMA codecs show better robustness than GSM ones.

### RxQual versus MOS

There are various researches on how some transmission parameters help evaluate the speech quality of a call. RxQual Full has been subject to several investigations in which a correlation with PESQ MOS has been suggested.
The RxQual tests results for clean speech and error conditions (Figure 5.18) demonstrate the following relation with AQuA PESQ MOS for each GSM codec: Figure 5.19 for FR, Figure 5.20 for EFR and, Figure 5.21 for HR. Also, each mark in the graphic curves has the noise level for which those MOS and RxQual values have been evaluated.

| Pn (dBm) | GSM FR | GSM EFR | GSM HR |
|---|---|---|---|
| no noise | | 0 | |
| -4,5 | 4 | 5 | 5 |
| -3,0 | 6 | 6 | 6 |
| -1,5 | 6 | 6 | 6 |
| 0,5 | 6 | 6 | 6 |
| 2,0 | 7 | 6 | 7 |
| 3,0 | 7 | 7 | 7 |

**Figure 5.18:** *RxQual in GSM (Clean Speech + Noise Levels)*

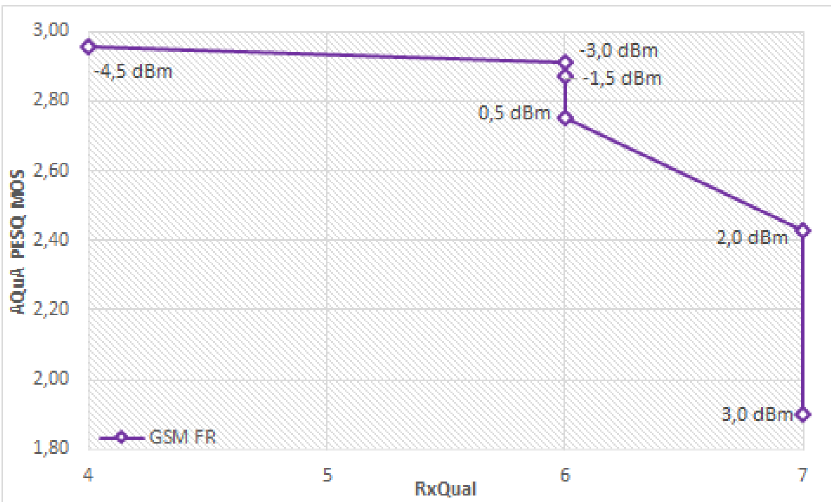



**Figure 5.19:** *RxQual versus AQuA PESQ MOS - GSM FR codec*

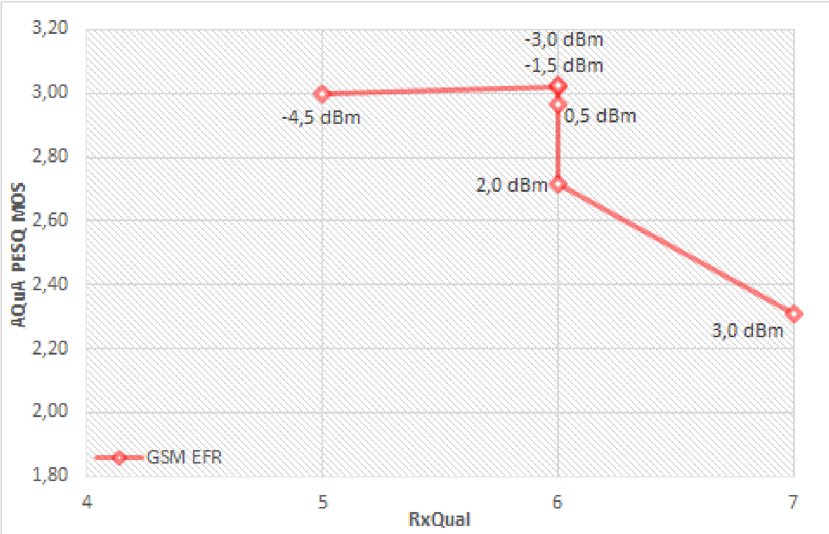



**Figure 5.20:** *RxQual versus AQuA PESQ MOS - GSM EFR codec*

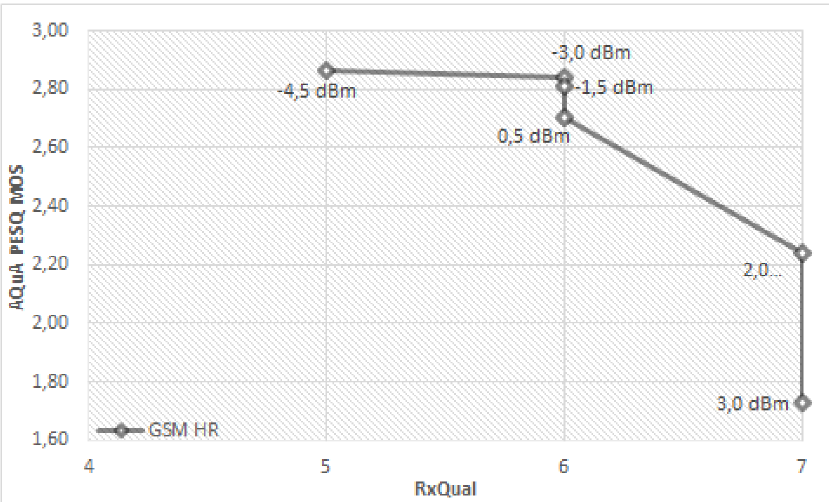



**Figure 5.21:** *RxQual versus AQuA PESQ MOS - GSM HR codec*

It is possible to assert from these measurements that speech quality only started to really deteriorate with RxQual 6, which coincides with Pn = 0.5 dBm of noise. The results also show that until RxQual 4 no significant change in the quality is perceived, thus verifying what is stated in [31] (chapter 3.2.2). The results also demonstrate that it takes longer for the EFR tests, to reach the worst possible RxQual value, which is in line with the inferences taken in the previous chapter.

Furthermore, studies in FR and EFR codecs from [62], [31] (page 95), [63] and [64] correlate with the curves shown in the above figures. This way being confirmed that RxQual, although not being a direct way of assessing speech quality, it is a helpful tool in aiding with the evaluation of the transmission, correlating with the speech quality. Since RxQual is an easily accessible transmission parameter it is, with no doubt, a most valuable tool for any engineer who needs to evaluate the communication channel.

### $Ec/N_0$ versus MOS

In this section, attempts to find a useful correlation between $Ec/N_0$ and MOS are made. No studies have been found on the matter, however it is relevant that such an important transmission parameter should be tested.

The $Ec/N_0$ (UMTS) lower and upper value limits for each Pn can be consulted on Figure 5.22.

| Pn (dBm) | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|
| | Lower/Upper (dB) | | |
| No Noise | - 2,5 / - 2.0 | | |
| - 4,5 | - 5,5 / -5.0 | - 4.0 / - 3,5 | - 5,5 / - 5.0 |
| - 3,0 | - 6.0 / - 5,5 | - 6.0 / - 5,5 | - 5,5 / - 5.0 |
| - 1,5 | - 7,5 / - 7.0 | - 7.0 / - 6,5 | - 6,5 / - 6.0 |
| 0,5 | - 9.0 / - 8,5 | - 7,5 / - 7.0 | - 9.0 / - 8,5 |
| 2,0 | - 10.0 / - 9,5 | - 10.0 / - 9,5 | - 10.0 / - 9,5 |
| 3,0 | - 11.0 / - 10,5 | - 10,5 / - 10.0 | - 11,5 / - 11.0 |

**Figure 5.22:** *$Ec/N_0$ (lower/upper values) in UMTS (Clean Speech + Noise Levels)*

Regarding $Ec/N_0$ values it is possible to verify that -2.0 dB is the best possible value achieved, even considering the clean speech condition.
For the correlation only the lower limit value of $Ec/N_0$ is considered.

From visualizing Figure 5.23 it is possible to state that speech quality values start decreasing significantly between -7.5 and -10 dB and, that until 7.5 dB the change in MOS is considered irrelevant. Also from -10 dB below, the speech quality exponentially decreases and, sometimes the call would drop.
Also, one might affirm that the behaviour of the UMTS tested codecs is quite similar from one to another in terms of $Ec/N_0$ thresholds.
Therefore it is concluded that there might be a sufficient correlation between $Ec/N_0$ and speech quality, worth of future studies.

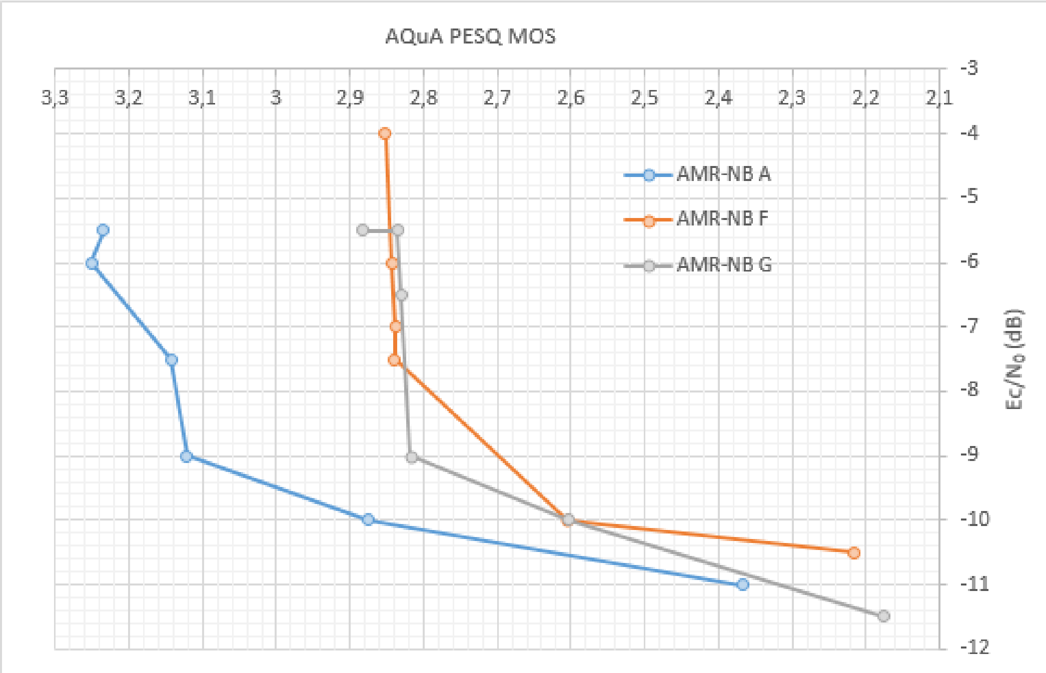



**Figure 5.23:** *$Ec/N_0$ versus AQUA PESQ MOS (UMTS codecs)*

### 5.2.6 Results Discussion

The limitations that resulted in non-optimal MOS values can be mainly justified by the following:

- Lack of proper Hardware/Software combination for the audio analysing. Which poses as the major setback in the experiment;

- On the account that an antenna was used instead of a wired connection (no external antenna connection cables were available for the phone), the duplex filter (diplexer) for use with the antenna was still missing. With which better filtration of the surrounding radio frequency emissions would have been possible.

Apart from this, one must always consider that in an experimental work, the results will always depend on the conditions in which the experiments were performed. So, results could vary, for example, depending on the phone under test (which could perform better for one cellular technology than the other), on the signal generator, or even in details like the loss in the conversion of the sample's bit rates.

Besides all the possible limitations of an experimental work, it can be concluded that the defined goals for this project were achieved with satisfactory results.

The overall codecs performance followed the expected behaviour (explained in 'MOS versus RF Noise' section). Also, a comparison and subsequent analysis of similar bit rate codecs between different technologies was successfully executed. Being able to conclude from this experiment that 3G speech codecs tested perform better whether in high or low bit rates, when compared with 2G speech codecs. With UMTS codecs reflecting greater robustness when RF noise is increased and also a higher quality (most easily verified in high rate codecs), even in similar coding based algorithms for the codecs.

Furthermore, transmission parameters RxQual and $Ec/N_0$ were evaluated in consonance with speech quality values. Confirming RxQual, as seen in other research studies, as a quite reliable speech quality indicator (only as an indicator, and not to evaluate speech quality by itself).

$Ec/N_0$ results show a good correlation with speech quality, not as a speech quality indicator but in another way: by indicating level thresholds for decreases in quality. A first $Ec/N_0$ level where a slight decrease of quality is noticed and then another level from which the speech quality degrades exponentially.

# 6

# Conclusions and Future Work

This dissertation allowed the acquiring of new competences and deeper knowledge in the area of mobile communications and telephonometry.
Through the study of the most up to date official recommendations from ITU, ETSI and 3GPP, along with recognized literature, it is possible to state that the foundations for the development of this work were firmly pre-established.

First, the study aimed to reinforce some generalised knowledge on the speech quality performance of 2G and 3G speech codecs and, at the same time, validating the system architecture used. On a second phase, further analysis were to be performed in order to better understand differences in the speech quality performance of equivalent bit rate codecs between the two cellular technologies.
Lastly, two transmission parameters were selected to verify their correlation with speech quality, with the purpose of validating them as relevant speech quality indicators. Concerning RxQual, previous research studies were utilized as a mean of validation. As for $Ec/N_0$, no prior studies were found on the matter therefore, new data was expected.

Even though the lack of proper hardware audio analyser for the experiment restricted the achievement of optimal measured results, with the conclusion of this work it is possible to assert that all the proposed objectives were accomplished with satisfactory levels.

The work's results allow for the following conclusions:

- The performance of the tested codecs, was behaviourally in conformance with the expected;
- In similar bit rates, 3G speech codecs performed better in terms of speech quality throughout all clean speech and error conditions imposed, in comparison with 2G. Proving to be more robust and with considerably better results in high bit rate codecs. Even when comparing similar based coding algorithms for the codecs in test, 3G still presented superior results;
- RxQual can be considered a quite reliable speech quality indicator (SQI), confirming the studies in [62], [31], [63] and [64]. Even though this parameter cannot alone be used to determine speech quality, the correlation between RxQual and MOS values demonstrates it as a good indicator of quality;
- Considering the $Ec/N_0$ results, it is not possible to state the parameter as a speech quality indicator, however, it shows clear thresholds for which the MOS values decrease significantly. A first $Ec/N_0$ level where a slight decrease of quality is noticed and then another level from which the speech quality degrades exponentially.

These results reinforce the idea that 3G is, with no doubt, the best choice if the costumer looks for the best possible listening speech quality. Adding to the fact that 3G is by far a better solution in the data transmission plane, when compared to 2G variants.
Also, the studied transmission parameters show that they can be used not only for network management purposes but at the same time give an expected idea to the communications engineer (or technician) of the end-to-end speech quality consequences.

Lastly, it is important that future studies should be carried on, giving continuity to this work.
Considering that the fourth-generation cellular technologies are now beginning to take an important place in the global market, as the first all-IP network structure, it seems of great relevance that 4G speech quality should be subject of evaluation. Comparing it to 3G, not only in narrowband but also adding wideband scenarios with the most recent standard objective method of speech quality assessment, POLQA. Also, further research on $Ec/N_0$ should be performed with the intention of further validating the assumptions made in this work.

# Bibliography

[1] J. Bannister, P. Mather, and S. Coope, in *Convergence Technologies for 3G Networks - IP, UMTS, EGPRS and ATM*, 2nd ed., L. John Wiley & Sons, ed., (2004).

[2] A. Perry, in *Fundamentals of Voice-Quality Engineering in Wireless Networks*, C. U. Press, ed., (2006).

[3] ITU-T, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Recommendation P.862, International Telecommunication Union (2001) .

[4] ITU-T, "Perceptual objective listening quality assessment," Recommendation P.863, International Telecommunication Union (2011) .

[5] S. Oy, "PESQ, POLQA, AQuA - Feature Comparison Tist,", `http://blog.sevana.fi/pesq-polqa-aqua/`, accessed on the 15th of June, 2013.

[6] E. Domiczi, "End-to-end speech and audio quality evaluation of networks using AQuA - competitive alternative for PESQ (P.862),", sevana Oy.

[7] J. Eberspacher, H. Vogel, C. Bettstetter, and C. Hartmann, in *GSM Architecture, Protocols and Services*, 3rd ed., L. John Wiley & Sons, ed., (2009).

[8] "R&S CMW280/500 Radio Communication Testers: GSM RF Testing - Features and Functions," Technical paper, Rohde & Schwarz (2012) .

[9] "R&S CMW280/500 Radio Communication Testers: WCDMA RF Testing - Features and Functions," Technical paper, Rohde & Schwarz (2012) .

[10] "Gigabyte GSmart G1305 Boston,", `http://www.gsmarena.com/gigabyte_gsmart_g1305_boston-3201.php`, accessed on the 16th of June, 2013.

[11] "ULTRA-DI DI100 - Professional Battery/Phantom Powered DI-Box," User manual, Behringer (2013) .

[12] GSA, "Q4 2012 mobile subscriptions,", `http://www.gsacom.com/`, accessed on the 14th of June, 2013.

[13] ITU-T, "Terms and Definitions Related to Quality of Service and Network Performance Including Dependability," Recommendation E.800, International Telecommunication Union (1994) .

[14] A. Mishra, in *Cellular Technologies for Emerging Markets - 2G, 3G and Beyond*, L. John Wiley & Sons, ed., (2010).

[15] A. Molisch, in *Wireless Communications*, 2nd ed., L. John Wiley & Sons, ed., (2011).

[16] L. Hanzo, F. Somerville, and J. Woodard, in *Voice and Audio Compression for Wireless Communications*, 2nd ed., L. John Wiley & Sons, ed., (2007).

[17] ITU-T, "Methods for subjective determination of transmission quality," Recommendation P.800, International Telecommunication Union (1996) .

[18] I. Cotanis, "Understanding the Transition from PESQ to POLQA - An Ascom Network Testing White Paper," White paper, Ascom (2011) .

[19] "POLQA - Perceptual Objective Listening Quality Analysis," White paper, Opticom and SwissQual (2011) .

[20] Y. Gaoxiong and Z. Wei, "The Perceptual Objective Listening Quality Assessment algorithm in telecommunication: Introduction of ITU-T new metrics POLQA," In *Communications in China (ICCC), 2012 1st IEEE International Conference on*, pp. 351–355 (2012).

[21] O. Gerlach, "Next-Generation (3G/4G) Voice Quality Testing with POLQA," White paper, Rohde and Schwarz (2012) .

[22] S. Oy, "Automated sound signals quality estimation,", confidential Property.

[23] S. Oy, "Automatic Sound Signals Quality Estimation,", `http://www.sevana.fi/automatic_sound_signals_quality_estimation_2.php`, accessed on the 15th of June, 2013.

[24] T. Halonen, J. Romero, and J. Melero, in *GSM, GPRS and EDGE Performance - Evolution Towards 3G/UMTS*, 2nd ed., L. John Wiley & Sons, ed., (2003).

[25] MIT, "MIT 6.02 DRAFT Lecture Notes,", chapter 9.3 - Bit Errors.

[26] Z. Adeyemo and T. Rajo, "Bit Error Rate analysis for wireless links using adaptive combining diversity," Journal of Theoretical and Applied Information Technology (2010).

[27] T. Rappaport, in *Wireless Communications: Principles and Practice*, 2nd ed., L. John Wiley & Sons, ed., (2002).

[28] M. Rowe, "BER measurements reveal network health,", `http://www.tmworld.com/design/design-and-prototyping/4381984/BER-measurements-reveal-network-health`, accessed on the 15th of June, 2013.

[29] W. Agilent, “Block Error Ratio (BLER) Measurement Description (January 16th, 2009),”, http://wireless.agilent.com/rfcomms/refdocs/wcdma/wcdma_meas_wblerror_desc.html, accessed on the 15th of June, 2013.

[30] ETSI, “Radio subsystem link control,” Technical specification 45.008, ETSI, (2012) , version 11.2.0 Release 11.

[31] R. Martin, U. Heute, and C. Antweiler, in *Advances in Digital Speech Transmission*, L. John Wiley & Sons, ed., (2008).

[32] “Understanding GSM Transmitter and Receiver Measurements for Base Transceiver Stations and Their Components,” Application note 1312, Agilent Technologies (Wireless/GSM Solutions)) .

[33] B. Walke, P. Seidenberg, and M. Althoff, in *UMTS: The Fundamentals*, L. John Wiley & Sons, ed., (2003), chapter 2.4.

[34] ETSI, “Performance characterization of the Adaptive Multi-Rate (AMR) speech codec,” Technical report 26.975, ETSI, (2012) , version 11.0.0 Release 11.

[35] H. Holma and A. Toskala, in *WCDMA for UMTS - HSPA evolution and LTE*, fourth ed., L. John Wiley & Sons, ed., (2007), chapter 3.3.

[36] M. Sauter, in *From GSM to LTE: An Introduction to Mobile Networks and Mobile Broadband*, L. John Wiley & Sons, ed., (2011).

[37] ETSI, “Codecs - ETSI,”, http://www.etsi.org/index.php/technologies-clusters/technologies/mobile/codecs, accessed on the 14th of June, 2013.

[38] A. Neubauer, J. Freudenberger, and V. Kuhn, in *Coding Theory - Algorithms, Architectures, and Applications*, L. John Wiley & Sons, ed., (2007).

[39] ETSI, “Half rate speech transcoding,” Technical specification 46.020, ETSI, (2012) , version 11.0.0 Release 11.

[40] ETSI, “Full rate speech processing functions,” Technical specification 46.001, ETSI, (2012) , version 11.0.0 Release 11.

[41] ETSI, “Half rate speech processing functions,” Technical specification 46.002, ETSI, (2012) , version 11.0.0 Release 11.

[42] ETSI, “Enhanced Full Rate (EFR) speech processing functions - General description,” Technical specification 46.051, ETSI, (2012) , version 11.0.0 Release 11.

[43] ETSI, “Discontinuous Transmission (DTX) for full rate speech traffic channels,” Technical specification 46.031, ETSI, (2012) , version 11.0.0 Release 11.

[44] ETSI, “Discontinuous Transmission (DTX) for half Rate (EFR) speech traffic channels,” Technical specification 46.041, ETSI, (2012) , version 11.0.0 Release 11.

[45] ETSI, “Discontinuous Transmission (DTX) for Enhanced Full Rate (EFR) speech traffic channels,” Technical specification 46.081, ETSI, (2012) , version 11.0.0 Release 11.

[46] ETSI, “Voice Activity Detector (VAD) for full rate speech traffic channels,” Technical specification 46.032, ETSI, (2012) , version 11.0.0 Release 11.

[47] ETSI, “Voice Activity Detector (VAD) for half rate speech traffic channels,” Technical specification 46.042, ETSI, (2012) , version 11.0.0 Release 11.

[48] ETSI, “Voice Activity Detector (VAD) for Enhanced Full Rate (EFR) speech traffic channels,” Technical specification 46.082, ETSI, (2012) , version 11.0.0 Release 11.

[49] ETSI, “Comfort noise aspect for full rate speech traffic channels,” Technical specification 46.012, ETSI, (2012) , version 11.0.0 Release 11.

[50] ETSI, “Comfort noise aspects for the half rate speech traffic channels,” Technical specification 46.022, ETSI, (2012) , version 11.0.0 Release 11.

[51] ETSI, “Comfort noise aspects for Enhanced Full Rate (EFR) speech traffic channels,” Technical specification 46.062, ETSI, (2012) , version 11.0.0 Release 11.

[52] D. Bruhn, E. Ekudden, and K. Hellwig, “Adaptive Multi-rate: A New Speech Service for GSM and Beyond,” Itg technical report 159, (2000) , proceedings 3rd ITG Conference Source and Channel Coding.

[53] ETSI, “AMR speech Codec - General description,” Technical report 26.071, ETSI, (2012) , version 11.0.0 Release 11.

[54] ETSI, “AMR speech Codec - Transcoding functions,” Technical specification 26.090, ETSI, (2012) , version 11.0.0 Release 11.

[55] ETSI, “Adaptive Multi-Rate - Wideband (AMR-WB) speech codec - Transcoding functions,” Technical specification 26.190, ETSI, (2012) , version 11.0.0 Release 11.

[56] ETSI, “AMR speech codec frame structure,” Technical specification 26.101, ETSI, (2012) , version 11.0.0 Release 11.

[57] ETSI, “AMR-WB speech codec frame structure,” Technical specification 26.201, ETSI, (2012) , version 11.0.0 Release 11.

[58] ITU-T, “Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB),” Recommendation g.722.2, ITU-T (2003) .

[59] ETSI, "AMR-WB speech codec - General description," Technical specification 26.171, ETSI, (2012) , version 11.0.0 Release 11.

[60] A. Ramo and H. Toukomaa, "On comparing speech quality of various narrow- and wideband speech codecs," In *Signal Processing and Its Applications, 2005. Proceedings of the Eighth International Symposium on*, **2,** 603–606 (2005).

[61] ETSI, "Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec: Transcoding functions," Technical specification 26.290, ETSI, (2012) , version 11.0.0 Release 11.

[62] B. Dolezalova, J. Holub, and M. Street, "Mobile network voice transmission quality estimation based on radio path parameters," In *Wireless Telecommunications Symposium, 2005*, pp. 95–99 (2005).

[63] H. Ming-Ju and A. Mostafa, "AMR Call Quality Measurement Based on ITU-T P.862.1 PESQ-LQO," In *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th*, pp. 1–5 (2006).

[64] M. Werner, K. Kamps, U. Tuisel, J. Beerends, and P. Vary, "Parameter-based speech quality measures for GSM," In *Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on*, **3,** 2611–2615 vol.3 (2003).

# Annex A. Test results - AQuA PESQ MOS

For AQuA PESQ MOS, results for each recording sample concerning the various noise levels (Pn) injected into the transmission channel, are presented in this Annex. From Figure 6.1 to Figure 6.6.

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| **F1** | 3,20 | 3,09 | 2,92 | 3,41 | 2,98 | 3,01 |
| | 3,20 | 3,08 | 2,92 | 3,59 | 3,01 | 2,99 |
| | 3,16 | 3,24 | 3,05 | 3,40 | 3,01 | 3,03 |
| **F3** | 2,98 | 3,38 | 3,02 | 3,15 | 3,04 | 3,06 |
| | 3,03 | 3,45 | 3,06 | 3,41 | 3,02 | 3,02 |
| | 3,03 | 3,03 | 3,04 | 3,32 | 3,02 | 3,01 |
| **M1** | 2,78 | 2,83 | 2,66 | 3,31 | 2,58 | 2,70 |
| | 2,86 | 2,87 | 2,66 | 2,88 | 2,66 | 2,73 |
| | 2,81 | 2,82 | 2,68 | 3,26 | 2,67 | 2,64 |
| **M3** | 2,71 | 2,72 | 2,77 | 2,99 | 2,71 | 2,84 |
| | 2,89 | 2,74 | 2,85 | 3,06 | 2,76 | 2,82 |
| | 2,85 | 2,76 | 2,74 | 3,03 | 2,76 | 2,74 |
| **Average Total** | **2,96** | **3,00** | **2,86** | **3,23** | **2,85** | **2,88** |
| **Standard Deviation** | **0,17** | **0,25** | **0,16** | **0,21** | **0,18** | **0,15** |
| **Average F1, F3** | 3,10 | 3,21 | 3,00 | 3,38 | 3,01 | 3,02 |
| **Average M1, M3** | 2,82 | 2,79 | 2,73 | 3,09 | 2,69 | 2,75 |

**Figure 6.1:** *AQuA PESQ MOS values (Pn = -4.5 dBm)*

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| **F1** | 3,11 | 3,11 | 2,94 | 3,24 | 3,03 | 2,82 |
| | 3,11 | 3,10 | 2,91 | 3,43 | 2,79 | 2,77 |
| | 3,18 | 2,98 | 3,03 | 3,35 | 3,09 | 3,08 |
| **F3** | 3,04 | 3,49 | 3,02 | 3,13 | 3,00 | 3,02 |
| | 3,07 | 3,40 | 2,84 | 3,65 | 3,01 | 2,99 |
| | 3,07 | 3,46 | 3,02 | 3,44 | 3,09 | 3,04 |
| **M1** | 2,81 | 2,66 | 2,56 | 3,24 | 2,63 | 2,61 |
| | 2,85 | 2,89 | 2,69 | 3,18 | 2,64 | 2,72 |
| | 2,81 | 2,75 | 2,54 | 3,17 | 2,68 | 2,72 |
| **M3** | 2,75 | 2,75 | 2,86 | 3,16 | 2,66 | 2,75 |
| | 2,73 | 2,86 | 2,83 | 2,99 | 2,77 | 2,69 |
| | 2,42 | 2,77 | 2,84 | 3,03 | 2,73 | 2,82 |
| **Average Total** | **2,91** | **3,02** | **2,84** | **3,25** | **2,84** | **2,84** |
| **Standard Deviation** | **0,22** | **0,29** | **0,17** | **0,19** | **0,19** | **0,16** |
| **Average F1, F3** | 3,10 | 3,26 | 2,96 | 3,37 | 3,00 | 2,95 |
| **Average M1, M3** | 2,73 | 2,78 | 2,72 | 3,13 | 2,69 | 2,72 |

**Figure 6.2:** *AQuA PESQ MOS values (Pn = -3.0 dBm)*

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| **F1** | 2,96 | 3,20 | 2,96 | 3,45 | 3,04 | 2,96 |
| | 3,00 | 3,13 | 2,95 | 3,35 | 2,75 | 3,06 |
| | 3,25 | 3,29 | 2,73 | 3,03 | 2,84 | 2,79 |
| **F3** | 2,99 | 3,35 | 2,99 | 3,15 | 2,94 | 2,98 |
| | 3,11 | 3,39 | 3,22 | 3,35 | 3,06 | 3,04 |
| | 2,99 | 3,45 | 3,05 | 3,14 | 3,10 | 3,01 |
| **M1** | 2,74 | 2,67 | 2,59 | 3,05 | 2,66 | 2,66 |
| | 2,73 | 2,99 | 2,62 | 3,23 | 2,61 | 2,68 |
| | 2,74 | 2,68 | 2,53 | 3,04 | 2,69 | 2,54 |
| **M3** | 2,44 | 2,66 | 2,82 | 3,03 | 2,81 | 2,73 |
| | 2,83 | 2,76 | 2,60 | 3,01 | 2,85 | 2,70 |
| | 2,66 | 2,72 | 2,67 | 2,88 | 2,71 | 2,81 |
| **Average Total** | **2,87** | **3,02** | **2,81** | **3,14** | **2,84** | **2,83** |
| **Standard Deviation** | **0,22** | **0,31** | **0,22** | **0,17** | **0,17** | **0,17** |
| **Average F1, F3** | 3,05 | 3,30 | 2,98 | 3,25 | 2,96 | 2,97 |
| **Average M1, M3** | 2,69 | 2,75 | 2,64 | 3,04 | 2,72 | 2,69 |

**Figure 6.3:** *AQuA PESQ MOS values (Pn = -1.5 dBm)*

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| **F1** | 2,83 | 3,15 | 2,76 | 3,28 | 2,80 | 2,82 |
| | 2,9 | 3,32 | 2,90 | 3,41 | 2,80 | 2,87 |
| | 2,06 | 3,18 | 2,63 | 3,04 | 2,82 | 2,77 |
| **F3** | 3,48 | 3,43 | 2,96 | 3,07 | 3,08 | 3,02 |
| | 2,77 | 3,02 | 2,87 | 3,52 | 3,42 | 2,96 |
| | 3,00 | 3,09 | 3,10 | 3,16 | 2,99 | 3,03 |
| **M1** | 2,58 | 2,70 | 2,50 | 3,10 | 2,68 | 2,71 |
| | 2,75 | 2,73 | 2,48 | 3,06 | 2,64 | 2,66 |
| | 2,71 | 2,91 | 2,45 | 2,94 | 2,63 | 2,69 |
| **M3** | 2,81 | 2,74 | 2,75 | 3,00 | 2,82 | 2,76 |
| | 2,78 | 2,75 | 2,43 | 2,98 | 2,70 | 2,72 |
| | 2,38 | 2,58 | 2,61 | 2,89 | 2,69 | 2,80 |
| **Average Total** | **2,75** | **2,97** | **2,70** | **3,12** | **2,84** | **2,82** |
| **Standard Deviation** | **0,34** | **0,27** | **0,22** | **0,19** | **0,23** | **0,13** |
| **Average F1, F3** | 2,84 | 3,20 | 2,87 | 3,25 | 2,99 | 2,91 |
| **Average M1, M3** | 2,67 | 2,74 | 2,54 | 3,00 | 2,69 | 2,72 |

**Figure 6.4:** *AQuA PESQ MOS values (Pn = -0.5 dBm)*

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| **F1** | 2,52 | 3,03 | 2,46 | 3,02 | 2,59 | 2,60 |
| | 2,51 | 2,78 | 2,31 | 2,83 | 2,58 | 2,43 |
| | 2,22 | 2,90 | 2,12 | 3,18 | 2,60 | 2,53 |
| **F3** | 2,64 | 3,28 | 2,39 | 2,93 | 2,82 | 2,83 |
| | 2,65 | 2,86 | 2,45 | 3,08 | 2,86 | 2,60 |
| | 2,48 | 2,98 | 2,17 | 2,99 | 2,77 | 3,17 |
| **M1** | 2,36 | 2,58 | 2,19 | 2,74 | 2,54 | 2,48 |
| | 2,55 | 2,53 | 2,11 | 2,92 | 2,47 | 2,45 |
| | 2,41 | 2,6 | 2,34 | 2,89 | 2,38 | 2,47 |
| **M3** | 2,28 | 2,12 | 2,09 | 2,54 | 2,52 | 2,54 |
| | 2,28 | 2,47 | 2,11 | 2,70 | 2,6 | 2,61 |
| | 2,23 | 2,43 | 2,15 | 2,68 | 2,54 | 2,52 |
| **Average Total** | **2,43** | **2,71** | **2,24** | **2,88** | **2,61** | **2,60** |
| **Standard Deviation** | **0,15** | **0,32** | **0,14** | **0,18** | **0,14** | **0,21** |
| **Average F1, F3** | 2,50 | 2,97 | 2,32 | 3,01 | 2,70 | 2,69 |
| **Average M1, M3** | 2,35 | 2,46 | 2,17 | 2,75 | 2,51 | 2,51 |

**Figure 6.5:** *AQuA PESQ MOS values (Pn = 2.0 dBm)*

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| **F1** | 1,91 | 2,67 | 1,99 | 2,43 | 2,40 | 1,98 |
| | 2,05 | 2,36 | 1,99 | 2,67 | 2,49 | 2,35 |
| | 2,09 | 2,37 | 1,26 | 2,51 | 2,45 | 2,41 |
| **F3** | 1,92 | 2,55 | 2,05 | 2,70 | 2,64 | 2,74 |
| | 1,60 | 2,60 | 1,39 | 2,43 | 2,61 | 2,40 |
| | 2,28 | 2,16 | 1,97 | 2,59 | 2,58 | 2,46 |
| **M1** | 2,03 | 2,25 | 1,84 | 2,26 | 2,17 | 2,03 |
| | 2,04 | 2,45 | 1,91 | 2,12 | 2,15 | 1,85 |
| | 2,07 | 2,36 | 1,89 | 2,16 | 1,13 | 2,28 |
| **M3** | 1,50 | 1,72 | 1,32 | 2,06 | 1,94 | 2,12 |
| | 1,73 | 2,12 | 1,34 | 2,23 | 2,00 | 1,47 |
| | 1,56 | 2,10 | 1,83 | 2,25 | 2,04 | 2,01 |
| **Average Total** | **1,90** | **2,31** | **1,73** | **2,37** | **2,22** | **2,18** |
| **Standard Deviation** | **0,25** | **0,26** | **0,31** | **0,22** | **0,42** | **0,34** |
| **Average F1, F3** | 1,98 | 2,45 | 1,78 | 2,56 | 2,53 | 2,39 |
| **Average M1, M3** | 1,82 | 2,17 | 1,69 | 2,18 | 1,91 | 1,96 |

**Figure 6.6:** *AQuA PESQ MOS values (Pn = 3.0 dBm)*

# Annex B. Test results - AQuA Voice Quality in Percentage

For AQuA Voice Quality in Percentage, results for each recording sample concerning the various noise levels (Pn) injected into the transmission channel, are presented in this Annex. From Figure 6.7 to Figure 6.12.

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| F1 | 83,74 | 81,29 | 77,29 | 88,42 | 78,60 | 79,35 |
| | 83,79 | 81,04 | 77,29 | 92,56 | 79,43 | 78,93 |
| | 82,88 | 84,76 | 80,40 | 88,29 | 79,36 | 79,74 |
| F3 | 78,63 | 87,86 | 79,67 | 82,62 | 80,06 | 80,64 |
| | 79,75 | 89,40 | 80,54 | 88,52 | 79,60 | 79,69 |
| | 79,74 | 79,92 | 80,03 | 86,46 | 79,55 | 79,39 |
| M1 | 73,82 | 75,03 | 70,71 | 86,26 | 68,64 | 71,70 |
| | 75,68 | 76,08 | 70,73 | 83,30 | 70,76 | 72,38 |
| | 74,54 | 74,82 | 71,09 | 85,02 | 70,85 | 70,18 |
| M3 | 72,02 | 72,23 | 73,45 | 78,92 | 72,07 | 75,25 |
| | 76,45 | 72,65 | 75,51 | 80,61 | 73,13 | 74,77 |
| | 75,42 | 73,15 | 72,86 | 79,88 | 73,25 | 72,65 |
| **Average Total** | **78,04** | **79,02** | **75,80** | **85,07** | **75,44** | **76,22** |
| **Standard Deviation** | **4,00** | **5,97** | **3,92** | **4,12** | **4,34** | **3,80** |
| **Average F1, F3** | 81,42 | 84,05 | 79,20 | 87,81 | 79,43 | 79,62 |
| **Average M1, M3** | 74,66 | 73,99 | 72,39 | 82,33 | 71,45 | 72,82 |

**Figure 6.7:** *AQuA Voice Quality in Percentage (Pn = -4.5 dBm)*

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| F1 | 81,64 | 81,73 | 77,73 | 84,66 | 79,95 | 74,74 |
| | 81,75 | 81,59 | 77,02 | 88,82 | 74,08 | 73,49 |
| | 83,22 | 78,82 | 79,85 | 87,14 | 81,21 | 81,09 |
| F3 | 80,18 | 90,31 | 79,64 | 82,12 | 79,21 | 79,69 |
| | 80,84 | 88,30 | 75,35 | 94,00 | 79,49 | 78,82 |
| | 80,87 | 89,65 | 79,60 | 89,09 | 81,13 | 80,10 |
| M1 | 74,53 | 70,69 | 57,98 | 84,68 | 69,82 | 69,37 |
| | 75,57 | 76,52 | 71,55 | 83,36 | 70,17 | 72,14 |
| | 74,59 | 73,08 | 67,46 | 83,07 | 71,08 | 72,20 |
| M3 | 72,94 | 73,03 | 75,82 | 92,76 | 70,74 | 73,01 |
| | 72,52 | 75,87 | 75,12 | 78,88 | 73,55 | 71,49 |
| | 64,27 | 73,43 | 75,27 | 79,86 | 72,50 | 74,81 |
| **Average Total** | **76,91** | **79,42** | **74,37** | **85,70** | **75,24** | **75,08** |
| **Standard Deviation** | **5,52** | **6,92** | **6,27** | **4,77** | **4,58** | **3,88** |
| **Average F1, F3** | 81,42 | 85,07 | 78,20 | 87,64 | 79,18 | 77,99 |
| **Average M1, M3** | 72,40 | 73,77 | 70,53 | 83,77 | 71,31 | 72,17 |

**Figure 6.8:** *AQuA Voice Quality in Percentage (Pn = -3.0 dBm)*

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| F1 | 78,29 | 83,82 | 78,10 | 89,31 | 80,00 | 78,24 |
| | 79,25 | 82,27 | 78,03 | 87,21 | 73,06 | 80,66 |
| | 84,78 | 85,70 | 72,54 | 79,75 | 75,31 | 74,02 |
| F3 | 78,81 | 87,02 | 78,91 | 82,73 | 77,83 | 78,73 |
| | 81,72 | 88,11 | 84,12 | 87,21 | 80,65 | 79,99 |
| | 78,86 | 89,28 | 80,33 | 82,48 | 81,53 | 79,29 |
| M1 | 72,64 | 71,04 | 68,80 | 80,22 | 70,72 | 70,65 |
| | 72,38 | 78,86 | 69,50 | 84,52 | 69,24 | 71,19 |
| | 72,65 | 71,26 | 67,23 | 80,16 | 71,35 | 67,47 |
| M3 | 64,71 | 70,59 | 74,85 | 79,78 | 74,45 | 72,54 |
| | 75,02 | 73,30 | 69,06 | 79,27 | 75,42 | 71,78 |
| | 70,59 | 72,32 | 70,85 | 76,30 | 72,07 | 74,63 |
| **Average Total** | **75,81** | **79,46** | **74,36** | **82,41** | **75,14** | **74,93** |
| **Standard Deviation** | **5,51** | **7,39** | **5,46** | **3,92** | **4,10** | **4,34** |
| **Average F1, F3** | 80,29 | 86,03 | 78,67 | 84,78 | 78,06 | 78,49 |
| **Average M1, M3** | 71,33 | 72,90 | 70,05 | 80,04 | 72,21 | 71,38 |

**Figure 6.9:** *AQuA Voice Quality in Percentage (Pn = -1.5 dBm)*

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| F1 | 75,13 | 82,64 | 73,20 | 85,60 | 74,37 | 74,70 |
| | 76,76 | 86,46 | 76,64 | 88,55 | 74,37 | 75,95 |
| | 54,02 | 83,33 | 69,96 | 80,10 | 74,79 | 73,56 |
| F3 | 90,14 | 88,92 | 78,15 | 80,74 | 81,10 | 79,57 |
| | 73,47 | 79,69 | 76,09 | 90,96 | 88,58 | 78,30 |
| | 79,19 | 81,33 | 81,54 | 92,96 | 78,90 | 79,83 |
| M1 | 68,48 | 71,76 | 66,39 | 81,57 | 71,16 | 72,07 |
| | 72,91 | 72,55 | 65,72 | 80,53 | 70,04 | 70,67 |
| | 71,99 | 76,93 | 65,11 | 77,67 | 69,94 | 71,42 |
| M3 | 74,44 | 72,78 | 73,02 | 79,19 | 74,79 | 73,21 |
| | 73,88 | 72,99 | 64,49 | 78,55 | 71,59 | 72,16 |
| | 63,10 | 68,57 | 69,29 | 76,46 | 71,54 | 74,20 |
| **Average Total** | **72,79** | **78,16** | **71,63** | **82,74** | **75,10** | **74,64** |
| **Standard Deviation** | **8,70** | **6,50** | **5,67** | **5,45** | **5,45** | **3,14** |
| **Average F1, F3** | 74,79 | 83,73 | 75,93 | 86,49 | 78,69 | 76,99 |
| **Average M1, M3** | 70,80 | 72,60 | 67,34 | 79,00 | 71,51 | 72,29 |

**Figure 6.10:** *AQuA Voice Quality in Percentage (Pn = -0.5 dBm)*

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| F1 | 67,16 | 79,39 | 65,63 | 79,32 | 68,89 | 69,18 |
| | 66,87 | 73,57 | 61,49 | 74,77 | 68,75 | 64,97 |
| | 59,21 | 76,51 | 56,39 | 82,82 | 69,15 | 67,58 |
| F3 | 70,35 | 85,11 | 63,82 | 77,27 | 74,64 | 74,78 |
| | 70,44 | 75,62 | 65,38 | 80,63 | 75,52 | 75,13 |
| | 66,28 | 78,21 | 57,84 | 78,62 | 73,37 | 82,51 |
| M1 | 62,94 | 68,86 | 58,40 | 72,74 | 67,73 | 66,22 |
| | 68,01 | 67,45 | 56,18 | 76,87 | 65,80 | 65,47 |
| | 64,31 | 69,18 | 62,39 | 76,24 | 63,61 | 65,96 |
| M3 | 60,89 | 56,48 | 55,39 | 67,64 | 67,31 | 67,73 |
| | 60,90 | 65,94 | 56,20 | 71,80 | 69,37 | 69,59 |
| | 59,33 | 64,85 | 57,20 | 71,13 | 67,79 | 67,19 |
| **Average Total** | **64,72** | **71,76** | **59,69** | **75,82** | **69,33** | **69,69** |
| **Standard Deviation** | **4,05** | **7,79** | **3,82** | **4,38** | **3,54** | **5,23** |
| **Average F1, F3** | 66,72 | 78,07 | 61,76 | 78,91 | 71,72 | 72,36 |
| **Average M1, M3** | 62,73 | 65,46 | 57,63 | 72,74 | 66,94 | 67,03 |

**Figure 6.11:** *AQuA Voice Quality in Percentage (Pn = 2.0 dBm)*

| | FR | EFR | HR | AMR-NB A | AMR-NB F | AMR-NB G |
|---|---|---|---|---|---|---|
| F1 | 51,00 | 70,80 | 53,40 | 64,94 | 64,14 | 53,17 |
| | 55,05 | 63,14 | 53,39 | 70,71 | 66,49 | 62,97 |
| | 56,14 | 63,61 | 33,76 | 67,00 | 65,56 | 64,6 |
| F3 | 51,36 | 68,02 | 54,89 | 71,59 | 69,98 | 72,29 |
| | 42,42 | 69,05 | 36,85 | 64,95 | 69,47 | 64,22 |
| | 61,13 | 70,50 | 52,87 | 68,79 | 68,66 | 65,66 |
| M1 | 54,39 | 60,48 | 48,97 | 60,69 | 58,37 | 54,35 |
| | 54,60 | 65,45 | 50,98 | 57,02 | 57,83 | 49,47 |
| | 55,66 | 63,31 | 50,60 | 58,12 | 57,28 | 61,27 |
| M3 | 39,70 | 45,74 | 35,18 | 55,35 | 52,00 | 57,68 |
| | 45,97 | 56,80 | 35,75 | 60,77 | 53,65 | 39,03 |
| | 41,35 | 56,44 | 48,90 | 60,45 | 54,69 | 53,77 |
| **Average Total** | **50,73** | **62,78** | **46,30** | **63,37** | **61,51** | **58,21** |
| **Standard Deviation** | **6,81** | **7,20** | **8,28** | **5,42** | **6,57** | **8,88** |
| **Average F1, F3** | 52,85 | 67,52 | 47,53 | 68,00 | 67,38 | 63,82 |
| **Average M1, M3** | 48,61 | 58,04 | 45,06 | 58,73 | 55,64 | 52,60 |

**Figure 6.12:** *AQuA Voice Quality in Percentage (Pn = 3.0 dBm)*